

Development of WWP2 ubiquitin ligase inhibitors using biophysical, structural and synthetic approaches.



Gregory Richard Hughes

School of Chemistry

University of East Anglia

This thesis is submitted for the degree of

Doctor of Philosophy

September 2019

© This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived therefrom must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

Acknowledgements

I would first like to thank Dr Andrew Chantry for his help and guidance over the past four years. I thank Dr Andrew Hemmings for introducing me to the dark art of protein crystallography. For the chemistry, I thank Professor Phil Page (I miss the parties), Dr Richard Stephenson and Dr Tom Storr. I would like to make a special thanks to the NMR group of Dr Jesus Angulo, Dr Sam Walpole and Dr Serena Monaco, who's collaboration in this project really helped to steer it in the right direction. I thank Dr Clare Stevenson for her help and assistance in the SPR. For my time at Charnwood, I thank Dr James Hitchin, who's experience and expertise in medicinal chemistry was particularly useful. I thank my mentor and friend, Dr Elvis Maduli. I would also like to give thanks to Carlos Lence, for welcoming me into his home for three months. I have met and worked with many friends over the past four years who I have made my time enjoyable. Dr Jess Watt and Dr Tiffany Yim, who first showed me the ways of working in a biology lab. (Dr to be) Adam Wright, for persuading me to do Cabaret 2018 ("It will be easy..."). I give a special mention to those who there from the very start; (Dr to be) Tom Bridge, Dr Ross Goodyear, Dr Gerald Keil and (Dr to be) Gareth Hughes. Also give thanks to members of the Page lab and third floor. Finally, I would like to thank my family and friends for their support, and a special mention to Alison.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements.

Gregory Richard Hughes

September 2019

Abstract

The ubiquitin- proteasome pathway (UPS) involves the targeting of proteins for their degradation by the 26S proteasome. Ubiquitin is covalently attached to targeted proteins via a three step enzymatic cascade. The three enzymes involved in this process are known as E1 activating, E2 conjugating and E3 ligases. The key to substrate selectivity is within the E3 ligases. E3 ligases can be divided into three domain containing groups: HECT, RING and RING between RING (RBR). Of particular interest is the Nedd4 group of HECT ligases. The Nedd4 group all share similar domain make up where they have substrate recruiting WW domains and a catalytic HECT domain, responsible for transferring ubiquitin onto the substrates. The mis-expression of a member of this group, WWP2, has been linked to tumour growth in various cancers. Targeting the UPS for cancer therapeutic use has already been shown to be achievable with the proteasome inhibitor Bortezomib. Targeting the more specific WWP2 ubiquitin ligase could lead to less side effects and greater efficacy. Research into developing inhibitors for WWP2 could lead to the discovery of novel therapeutics.

The aim of this thesis is to design, synthesise and evaluate novel WWP2 ubiquitin ligase inhibitors. Different methodologies are explored in order to reveal WWP2 small molecule binding events. A thermal shift screen is employed to search for binders, and orthogonal methods for validation are tested. Using these methods, novel WWP2 ligands are discovered with inhibition activity. Crystallography is used to structurally elucidate a novel binding site for WWP2 ligands. Data acquired from ligand NMR techniques is used to validate this binding site. The combined X-ray crystallography and ligand NMR data is used to inform organic

synthesis of analogues for a particular inhibitor. These analogues are tested for inhibition against WWP2 and Nedd4 autoubiquitination. IC₅₀ values generated from this analogue series further validate the novel binding site and also offer higher levels of activity of below 1 μ M.

Table of contents

List of Figures	xv
List of tables	xix
1 Introduction	1
1.1 Ubiquitination	3
1.1.1 Ubiquitin	5
1.1.2 E1 activating and E2 conjugating enzymes	6
1.1.3 E3 ligases	10
1.1.3.1 RING ligases	10
1.1.3.2 HECT ligases	13
1.1.3.3 Nedd4 E3 ligases	16
1.1.3.4 WWP2: Structure, function and role in cancer	18
1.2 Therapeutic targeting of the UPS	21
1.2.1 Proteasome inhibitors	21
1.2.2 E1 and E2 inhibitors	22
1.2.3 Targeting the E3 ligases	24
1.2.3.1 PROTACs	24
1.2.3.2 RING inhibitors	25
1.2.3.3 HECT inhibitors	25

1.2.3.4	Other HECT ligase inhibitors	31
1.3	Thesis aims	32
2	Biological materials and methods	33
2.1	DNA techniques	34
2.1.1	Plasmids and primers	34
2.1.2	Phusion PCR	35
2.1.3	DNA Gel electrophoresis	35
2.1.4	Plasmid digest	36
2.1.5	In-Fusion ligation	36
2.1.6	DNA extraction and purification	37
2.1.7	DNA transformations	37
2.1.8	Plasmid purification	37
2.2	Protein techniques	37
2.2.1	Protein expression	37
2.2.2	Protein purification	38
2.2.2.1	His tagged proteins	38
2.2.2.2	GST tagged protein	38
2.2.2.3	Dialysis	39
2.2.2.4	GST cleavage	39
2.2.2.5	Gel filtration	39
2.2.3	Protein crystallisation	40
2.2.3.1	Crystal formation	40
2.2.3.2	Crystal harvesting	40
2.2.3.3	X-ray data collection	40
2.2.4	SDS-PAGE gel electrophoresis	41
2.2.5	Plate coating	42

2.2.6	Autoubiquitination assay	42
2.2.7	Differential scanning fluorimetry	43
2.2.8	Hydrogen peroxide detection assay	43
2.2.9	Surface plasmon resonance	44
2.2.10	Statistics and repeats	44

3 Efforts towards the discovery and development of WWP2 inhibitors through biophysical and synthetic methods 45

3.1	Introduction	46
3.1.1	Biophysical techniques	47
3.1.1.1	Differential scanning fluorimetry	47
3.1.1.2	Surface plasmon resonance	48
3.1.1.3	Saturation transfer difference NMR	48
3.1.2	Experimental aims	50
3.2	Results	51
3.2.1	Utilising a thermal shift assay to identify WWP2 ligands	51
3.2.1.1	WWP2 HECT protein purification	51
3.2.1.2	Thermal shift buffer screen	52
3.2.1.3	Compound screen	53
3.2.1.4	Biological testing of hit compounds	56
3.2.2	Orthogonal approaches to validate ligands	60
3.2.2.1	Orthogonal validation by STD NMR	60
3.2.2.2	Using surface plasmon resonance to investigate binding to WWP2	62
3.2.2.3	Purification of WWP2 FL	62
3.2.2.4	WWP2 SPR experiments	63
3.2.2.5	Detection of ligand binding to WWP2 by SPR	64

3.2.3	Synthetic approaches	67
3.2.3.1	Heclin synthesis	67
3.2.3.2	Vilsmeier-Haak reaction	68
3.2.3.3	Knoevenegal condensation Doebner modification	69
3.2.3.4	Amide bond formation	70
3.2.3.5	Synthesis of the purpurogallin scaffold	70
3.2.3.6	Purpurogallin analogue testing	74
3.3	Discussion	76
4	Towards the solving of the first ligand bound WWP2 crystallographic structure	80
4.1	Introduction	81
4.1.1	Crystal formation	81
4.1.2	X-ray diffraction	82
4.1.3	Protein ligand structures of the HECT ubiquitin ligases	83
4.1.4	Chapter aims	84
4.2	Results	85
4.2.1	Identification of putative binding sites	85
4.2.2	Attempting WWP2 HECT crystallisation	87
4.2.3	WWP2 WW2-2,3 linker HECT purification and crystallisation	89
4.2.4	Initial WWP2 crystal harvesting	90
4.2.5	Further soaking experiments leads to improved search model	93
4.2.6	First potential protein ligand structure	96
4.2.7	Optimising of soak experiments to reveal further protein ligand structures	102
4.3	Discussion	112
5	Utilising ligand NMR in the synthesis of isoalloxane based WWP2 inhibitors	116

5.1	Introduction	117
5.1.1	Discovery of novel WWP2 inhibitors by biochemical approaches	117
5.1.2	Ligand NMR	118
5.1.2.1	STD Epitope maps	118
5.1.2.2	DEEP STD NMR	119
5.1.3	Experimental aims	120
5.2	Results	121
5.2.1	<i>In silico</i> modelling of NSC288387 into WWP2 HECT	121
5.2.2	NMR validation of the WWP2:NSC288387 <i>in silico</i> model	123
5.2.3	<i>In silico</i> modelling of NSC288387 into WWP1 HECT	126
5.2.4	NMR validation of the WWP1:NSC288387 <i>in silico</i> model	126
5.2.5	<i>In silico</i> modelling of NSC288387 into Nedd4 HECT	129
5.2.6	Design of NSC288387 analogues	129
5.2.7	Organic synthesis	131
5.2.7.1	Amine protection and alkylation	131
5.2.7.2	Nucleophilic aromatic substitution	132
5.2.7.3	Reduction of nitro group and condensation with alloxan monohydrate	133
5.2.7.4	Chan- Lam coupling	134
5.2.8	Testing of analogues against WWP2 autoubiquitination	139
5.2.9	Testing of compounds against Nedd4	143
5.3	Discussion	147
6	Final Discussion	152
6.1	Introduction	153
6.2	Gaining a starting point	153
6.3	The PAIN argument	154

6.4	A novel allosteric binding site	157
6.5	Future work	158
7	Chemistry experimental	161
7.1	General chemistry experimental	162
7.1.1	Preparation of solvents and glassware	162
7.1.2	Compound characterisation	162
7.1.3	Chromatographic techniques	163
7.1.4	Compound numbering	163
7.2	Individual procedures and characterisation	164
	References	186

Abbreviations

AKT- Protein kinase B

AMP- Adenosine monophosphate

ATP- Adenosine triphosphate

BARD1- BCRA1 associated RING domain protein 1

BRCA1- Breast cancer type 1 susceptibility protein

BIRC7- Baculoviral IAP repeat containing protein 7

C2- Calcium binding domain

cCBL- Casitas B lineage lymphoma

DEEP- Differential epitope mapping

DMSO- Dimethylsulfoxide

DNA- Deoxyribonucleic acid

dNTP- Deoxynucleoside triphosphate

DSF- Differential scanning fluorimetry

DCM- Dichloromethane

DMF- N, N dimethylformamide

ELISA- Enzyme linked immunosorbent assay

EtOAc- Ethyl acetate

FL- Full length

GST- Glutathione s-transferase

HECT- Homologous to the E6AP carboxyl terminus

IPTG- Isopropyl- β -D- thiogalactoside
 ITCH- Itchy E3 ubiquitin protein ligase
 kDa- Kilodalton
 LB- Lysogeny broth
 LCMS- Liquid chromatograph mass spectrometry
 MES- 2-(N-morpholino) ethanesulfonic acid
 MMT- Malic acid, MES, Tris
 mRNA- Messenger RNA
 MW- Molecular weight
 Nedd4- Neural precursor cell expressed developmentally down regulated protein 4
 NMR- Nuclear magnetic resonance
 NSC- Cancer chemotherapy national service center number
 NOE- Nuclear overhauser effect
 PAGE- Polyacrylamide gel electrophoresis PCR- Polymerase chain reaction
 PDB- Protein data bank
 PROTAC- Proteolysis targeting chimera
 PTEN- Phosphatase and tensin homolog
 RBR- RING between RING
 RING- Really interesting new gene
 RMSD- Root mean square deviation
 RNA- Ribonucleic acid
 SDS- Sodium dodecyl sulfate
 SMAD- Small mothers against decapentaplegic
 SMURF- Smad ubiquitin regulatory factor
 S_NAr- Nucleophilic aromatic substitution
 SPR- Surface plasmon resonance

STD- Saturated transfer difference

TGF β - Transforming growth factor beta

THF- Tetrahydrofuran

TMB- Tetramethylbenzidine

tRNA- Transfer ribonucleic acid

UBA- Ubiquitin like modifier activating enzyme

UBC- Ubiquitin conjugating

UPS- Ubiquitin proteasome system

VHL- Von Hippel Lindau

WWP1- WW domain containing E3 ubiquitin protein ligase 1

WWP2- WW domain containing E3 ubiquitin protein ligase 2

List of Figures

1.1	Schematic of the ubiquitin-proteasome pathway	4
1.2	The structure of ubiquitin	5
1.3	Bound structure of UBA1 and two ubiquitin molecules.	7
1.4	Structure of E2 bound structures	9
1.5	Structure of RING complexes	12
1.6	Structural representation of the L and T shape conformation of HECT . . .	15
1.7	Schematic showing the Nedd4 superfamily of HECT E3 ligases.	17
1.8	Structure of WWP2 HECT domain	19
1.9	Genomic analysis of WWP2 and the highly homologous WWP1 in prostate cancer.	20
1.10	Selection of proteasome inhibitors	22
1.11	Small molecule inhibitors of E1 and E2.	23
1.12	Inhibitors of E3 RING ligases.	26
1.13	Inhibitors of E3 HECT ligases	27
1.14	Structure of the Nedd4 inhibitor	30
3.1	Nickel NTA purification of WWP2 HECT.	52
3.2	Condition screen for thermal shift	53
3.3	Thermal shift screen	55
3.4	Thermal shift hit compounds IC50's	59

3.5	STD spectrum of NSC35676	61
3.6	NMR spectrum showing the degradation of NSC401005 in the presence of WWP2 HECT	61
3.7	WWP2 FL protein purification	63
3.8	SPR sensorgram of WWP2 fl and WWP2 HECT	64
3.9	Compound binding to WWP2 HECT	65
3.10	R-max equation	65
3.11	R-max values of compound binding to WWP2 HECT	66
3.12	Scheme 1	68
3.13	Scheme 2	69
3.14	Scheme 3	70
3.15	Scheme 4	71
3.16	Scheme 5	73
3.17	Scheme 6.	73
3.18	Purpurogallin analogue testing	74
3.19	Purpurogallin analogue testing against Nedd4 autoubiquitination	75
3.20	HS-152 and NSC215718	78
3.21	Purpurogallin containing the catechol moiety	79
4.1	Nedd4 HECT bound to "Compound 1"	83
4.2	Predicted small molecule binding sites for WWP2 HECT.	86
4.3	WWP2 HECT purification for crystallography	88
4.4	Key steps in the WW2-2,3 linker HECT purification	90
4.5	PDB validation of new search model	97
4.6	WWP2_6_1 double difference Fourier electron density map.	98
4.7	Protein ligand interactions present in WWP2_6_1.	100
4.8	Double difference Fourier electron density of NSC2805 in WWP2_34_3. . .	105

4.9	WWP2_34_3 protein ligand interactions.	107
4.10	NSC2805 electron density in WWP2_38_1.	110
4.11	Electron density map comparison between apo (A) and soaked NSC2805 (B)	115
5.1	Schematic of the WWP2 autoubiquitination assay	118
5.2	Equations used for calculating ligand epitope maps	119
5.3	Equation used to calculate the DEEP STD factor	120
5.4	Structure of NSC288387	121
5.5	NSC288387 docked into WWP2 HECT.	123
5.6	Ligand epitope map of NSC288387 bound to WWP2 HECT	124
5.7	DEEP STD factors for 288387 bound to WWP2 HECT	126
5.8	Ligand epitope of NSC288387 when bound to WWP1 HECT	127
5.9	DEEP STD factors for 288387 bound to WWP1 HECT	128
5.10	Scheme 7	132
5.11	Scheme 8	132
5.12	Scheme 9	133
5.13	Scheme 10	135
5.14	A simplified Chan Lam catalytic cycle.	136
5.15	Scheme 11	137
5.16	Scheme 12	137
5.17	Scheme 13	137
5.18	Scheme 14	138
5.19	Scheme 15	138
5.20	WWP2 autoubiquitination IC50 curves	140
5.21	Nedd4 autoubiquitination IC50 curves	144
5.22	<i>In silico</i> structure of compound 21 docked into the binding site.	150
6.1	Phenol red assay.	156

6.2	Preliminary data showing the thermal shift screen using the WWP2 WW2-2,3	
	linker-HECT construct	159

List of tables

2.1	Plasmids used throughout this project	34
2.2	Primer sequences	34
2.3	PCR reaction mixture	35
2.4	PCR cycle conditions	35
2.5	Restriction enzyme digest reaction	36
2.6	Plasmid ligation reaction mixture	36
2.7	Protein expression conditions	38
2.8	Ni NTA purification buffer recipes	39
2.9	GST purification buffer recipies	40
2.10	10 % and 15 % acrylamide gel components and volumes for 2 gels	41
3.1	Structures and $\Delta T_m D$ of Diversity set hits	56
3.2	Structures and $\Delta T_m D$ of Natural product set hits	57
3.3	IC ₅₀ values of thermal shift hit compounds for the WWP2 autoubiquitination assay	58
4.1	Summary of first soaking experiments	92
4.2	Summary of the solved data from the first soaking experiments	92
4.3	Summary of second soaking experiments	94
4.4	Summary of second soaking experiments	95

4.5	Validation data comparison of the new search model and the published structure	96
4.6	WWP2_6_1_3 validation	101
4.7	Summary of glycerol titration soaking experiments	102
4.8	Summary data collected from glycerol titrations	102
4.9	Summary of the DMSO soaking experiments	102
4.10	Summary of the solved data from the DMSO soaking experiments	103
4.11	Summary of the NSC2805 soaking experiments in the improved soak conditions	103
4.12	Summary of the solved data from the improved soaking experiments	104
4.13	WWP2_34_3 validation	108
4.14	Summary of long soak experiments	109
4.15	Summary of the solved data from the long soaking experiments	109
4.16	WWP2_38_1 validation table	111
5.1	WWP2 HECT: NSC288387 STD epitope values	125
5.2	WWP2 DEEP STD factors	125
5.3	WWP1 HECT: NSC288387 STD epitope values.	127
5.4	WWP1 DEEP STD factors	128
5.5	Ranked analogues	131
5.6	Chan Lam substrate yields	135
5.7	NSC288387 analogue WWP2 autoubiquitination IC50's	142
5.8	NSC288387 analogue Nedd4 autoubiquitination IC50's	146

Chapter 1

Introduction

The central dogma of biology describes how genetic material of an organism codes for biological function. DNA is the language used to encode this information. The DNA sequence is converted to mRNA by RNA polymerase in a process known as transcription. The mRNA is then decoded by matching up with complementary tRNA. The tRNA molecules hold the specific amino acid which is coded for by the three lettered DNA sequence. The aligned amino acids join via a peptide bond to form a polypeptide chain. The final polypeptide chain is folded into the chosen structure, the protein. Proteins are nature's machines, armed with the ability to do a vast range of functions. Proteins can have enzymatic capability to catalyse reactions. They can be involved in controlling cell signalling pathways in order to produce the desired response. Proteins can also act as structural scaffolds to change the physical properties of tissue. Proteins can undergo post translational modifications which can in turn modify a protein's role. One such post translational modification is ubiquitination. Ubiquitination involves the labelling of proteins with a small protein named ubiquitin. The labelling of ubiquitin can have proteolytic and non proteolytic consequences. The attachment of ubiquitin is carried out by the E3 ubiquitin ligases. WWP2 (WW domain containing protein 2) is a HECT E3 ligase member of the Nedd4 family. The currently known role for WWP2 is to attach ubiquitin to the target substrates through its intrinsic activity, labelling the target for degradation by the proteasome. Mis-regulation of WWP2 has been linked to disease, prominently cancer. Cancer is globally one of the leading causes of death, approximately 1 in 6 deaths are associated with cancer (statistic from the World Health Organisation). Genetic mutations at the protein level can lead to cellular transformation, leading to uncontrolled cell growth which has damaging effects on the body. WWP2 targets the tumour suppressors Smads and pTEN in the TGF β pathway and PI3K/AKT pathway respectively.

This chapter introduces the process and functional roles of protein ubiquitination. A current insight regarding inhibitors of the ubiquitination pathway, mainly focusing on the E3 ligases are summarized.

1.1 Ubiquitination

Maintaining control over protein levels is key to cellular homeostasis. The rate at which proteins are synthesised and degraded determines protein turnover. In eukaryotic cells, protein degradation is conducted by two major pathways; lysosomal proteolysis and the ubiquitin proteasome pathway. Lysosomes are membrane enclosed organelles which hold a selection of digestive enzymes. Cellular proteins are taken up by the lysosome and degraded (Cooper et al., 2000). The ubiquitin proteasome pathway (Figure 1.1) provides a greater level of specificity in protein degradation, and will be discussed here in further detail. Ubiquitination is the sequential cascade of enzymatic events which labels proteins with ubiquitin. Ubiquitin is a small 76 amino acid protein found in most eukaryotic organisms. The seminal work by Ciechanover, Hershko and Rose discovered the role of ubiquitination in ATP activated protein turnover (Ciechanover et al., 1978) (Hershko et al., 1979) (Hershko et al., 1980). The pathway has later been found to be involved in other processes such as protein trafficking and DNA repair (Chantry, 2011). The three key enzymes involved in ubiquitination are known as E1 activating, E2 conjugating and E3 ligases. Ubiquitination begins with the ATP catalysed formation of a thio-ester bond between an E1 activating enzyme and ubiquitin. Ubiquitin is subsequently bound to an E2 conjugating enzyme which leads onto the formation of an isopeptide bond between the ubiquitin carboxyl terminus and a lysine residue of the target substrate. The process as to how the substrate is ubiquitinated from the E2 enzyme is mediated through the E3 ligases (Scheffner and Kumar, 2014). Target substrates can be mono-ubiquitinated or poly-ubiquitinated, and the latter marks the protein for degradation by the 26S proteasome. The 26S proteasome is the large multi component complex which degrades proteins via proteolysis.

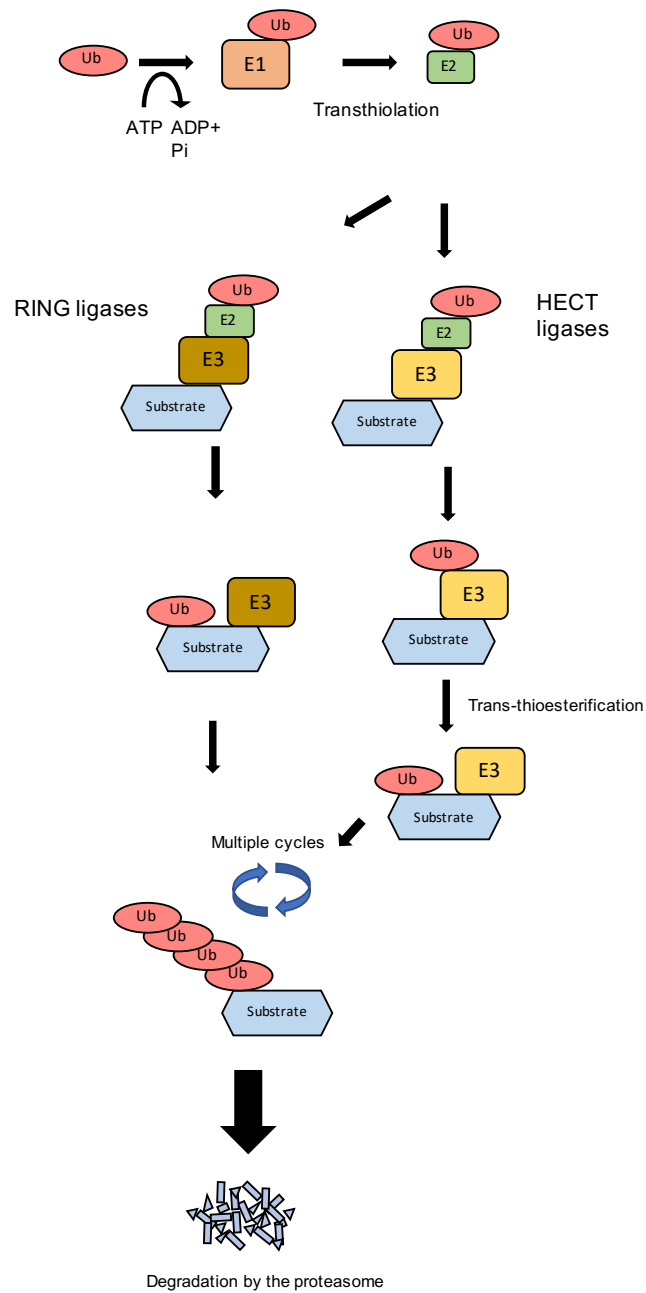


Fig. 1.1 -Schematic of the ubiquitin- proteasome pathway. The HECT and RING pathways are differentiated by the HECT's intrinsic ability to transfer ubiquitin, whilst the RING ligases mediate the transfer of ubiquitin from E2 to the substrate. Not to scale.

1.1.1 Ubiquitin

Discovered in 1975 (Goldstein et al., 1975), ubiquitin is a 8.6 kDa protein consisting of 76 amino acids. The tightly folded structure (Figure 1.2) contributes to a significant number of intramolecular hydrogen bonding. It is the high degree of hydrogen bonding which confers its stability over a broad range of pH and temperature values, and resistance to digestion (Vijay-Kumar et al., 1987). The carboxyl group on the C-terminal glycine of ubiquitin forms an isopeptide bond with the ϵ amino group of lysine on the substrate. The building of ubiquitin chains can occur at any one of the seven lysine residues on ubiquitin. The lysine residue chosen, and the degree of polymerisation dictates the fate of the substrate, which include DNA repair, kinase activation and proteasomal degradation (Sun and Chen, 2004).



Fig. 1.2 -The structure of ubiquitin. The predominant secondary structure of ubiquitin consists of an α - helix, a 5 stranded β - sheet and a 3_{10} helix. Other features include a parallel G1 β -bulge, two reverse Asx turns and a symmetrical hydrogen bonding region between two reverse turns and two helices. Structure solved by X-ray crystallography to 1.8 Å (Vijay-Kumar et al., 1987). PDB:1UBQ . Image generated by UCSF Chimera (Pettersen et al., 2004).

1.1.2 E1 activating and E2 conjugating enzymes

The ubiquitination cascade starts with the E1 activating enzyme. There are 8 E1 enzymes, but only 2 are known to interact with ubiquitin; Ubiquitin- like modifier activating enzyme 1 (UBA1) and ubiquitin- like modifier activating enzyme 6 (UBA6) (Schulman and Harper, 2009). Both UBA1 and UBA6 are known to initiate the ubiquitination pathway by the ATP dependent recruitment of ubiquitin. However, the several fold higher expression levels of UBA1 suggests the majority of ubiquitin pathways are dependent on UBA1 (Yang et al., 2013). UBA1 is a highly conserved protein which exists in two isoforms: UBA1a and UBA1b. The expression of UBA1 is essential and its deletion is fatal (Groen and Gillingwater, 2015). The E1 is a 110-120 kDa monomeric protein consisting of the two adenylation domains, the catalytic cysteine domains and the E2 recruitment domain. E1 initiates the ubiquitin cascade by catalysing the adenylation of the C-terminus of ubiquitin in an ATP dependent manner. The catalytic cysteine of the E1 then forms a thioester bond with the ubiquitin molecule. A second ubiquitin molecule will then attach to the adenylation site and remains in that intermediate state until thiolation can occur (Lee and Schindelin, 2008). This is shown structurally in Figure 1.3 with the two bound ubiquitin molecules on UBA1.

The second ubiquitin binding E1, UBA6 has been implicated in having overlap with UBA1 activity (Groettrup et al., 2008). Liu et al. have also recently showed UBA6 suppresses epithelial mesenchymal transition in mammary epithelial cells highlighting the potential for this gene to be used as a biomarker. There is a need for further investigation in the impact this E1 has on cellular regulation, UBA1 isoform interaction and any potential implications in disease. Following from the E1 enzymes in the ubiquitination cascade, are the E2-conjugating enzymes. There are around 40 human E2's presenting an increase in the complexity in the pathway. All E2s are characterised by the distinctive ubiquitin conjugating cysteine (UBC) domain which holds the catalytic cysteine. Structurally, the domain consists of 4 α -helices and a 4 stranded β - sheet. The E2 enzyme forms a non-covalent interaction with the E1-

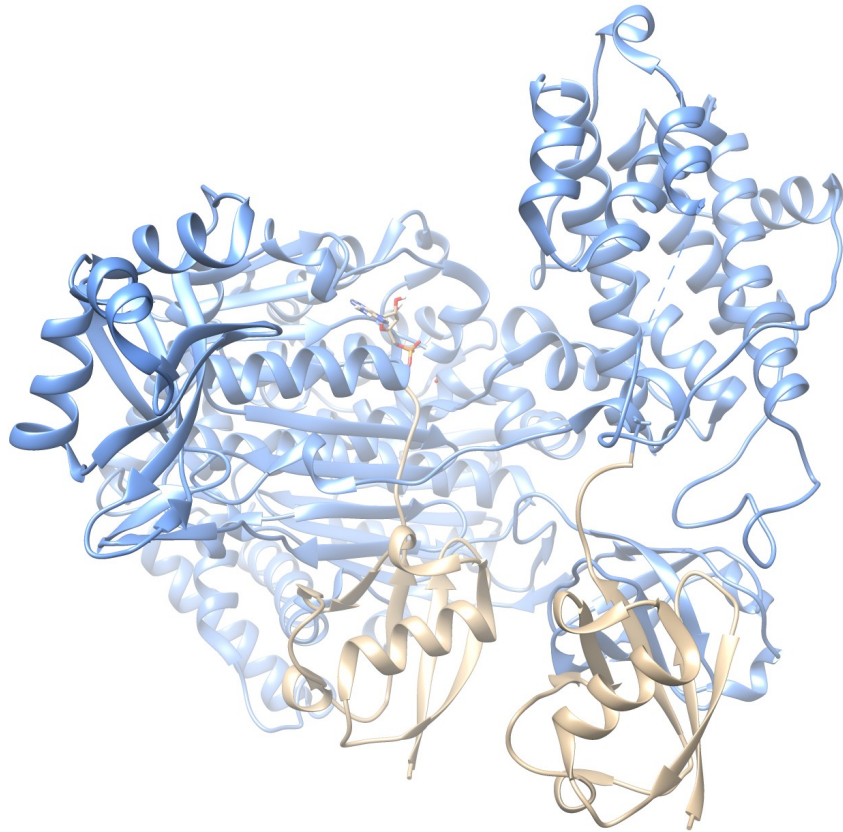


Fig. 1.3 -Bound structure of UBA1 and two ubiquitin molecules. Complex of *Saccharomyces Cerevisiae* UBA1 (blue) bound to ubiquitin-AMP and thioesterified ubiquitin. The ubiquitin on the left is in its adenylated form, located in the adenylation site. The ubiquitin on the right has formed the thioester bond with the catalytic cysteine of E1 (Cys600). Structure solved by X-ray crystallography (Schäfer et al., 2014). PDB: 3CMM. Image generated by UCSF Chimera (Pettersen et al., 2004).

Ub complex and the ubiquitin is transferred through a transthioylation reaction. Figure 1.4 shows the tertiary complex formed between E1, E2 and Ub. The E2-Ub complex will then interact with a specific E3 ligase to transfer the ubiquitin onto the target substrate. Not all E2s maintain functionality with all E3 ligases; the E2 UbCh7 will only show reactivity towards the HECT catalytic cysteine and not lysine residues presented by RING ligases. The implications of this is that whilst the UbCh7-Ub complex will interact with RING ligases, it will only function as an E2 with HECT ligases (Stewart et al., 2016). The E2 which is selected in a given pathway, determines the level of ubiquitination of the final substrate, whether that be mono or poly ubiquitination (Windheim et al., 2008). The control E2s have in the recruitment of E3s and the level of substrate ubiquitination, adds a layer of specificity in the ubiquitin pathway.

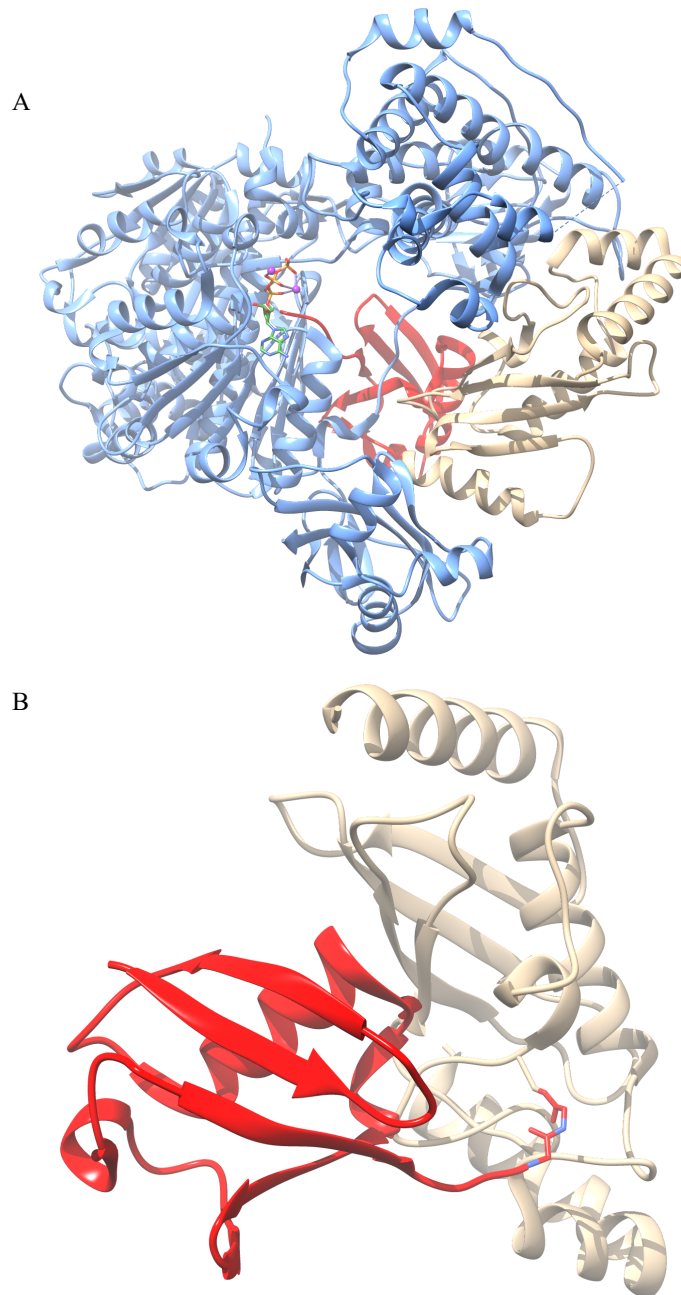


Fig. 1.4 -Structure of E2 bound structures (A) Structure of *Schizosaccharomyces Pombe* Uba1 (blue)-Ub (red)-Ubc4 (brown)/ ATP-Mg complex. Solved by x-ray crystallography (Olsen and Lima, 2013). PDB: 4II2. (B) Structure of bound UbcH5b (brown)- ubiquitin (red). The catalytic Cys85 was substituted to a serine residue as the thioester linked conjugate is unstable and unsuitable for crystallisation. Hence, the linkage is an ester bond. The structure was solved by X-ray crystallography (Sakata et al., 2010). PDB: 3A33. Image generated by UCSF Chimera (Pettersen et al., 2004).

1.1.3 E3 ligases

At the final step of the enzymatic cascade lies the E3 ligases. With over 600 human E3s, it is the E3 ligases which give the greatest level of diversity and highest level of specificity in the ubiquitin proteasome system (UPS). E3s can interact with E2s, ubiquitin and the target substrate to deliver the final ubiquitinated target. Based on their molecular architecture and mechanism of action, E3s can be divided into three classes: really interesting new gene (RING), homologous to the E6AP C-terminus (HECT) and RING between RING (RBR). RBR is the smallest of the three groups and the least investigated, therefore only RING and HECT ligases will be discussed in further detail.

1.1.3.1 RING ligases

With around 600 known members, the RING-type ligases constitute the majority of E3 ligases. RING type ligases can be more specifically divided into the RING and RING-like U-box families. RINGs adopt a cross-braced arrangement with eight centrally co-ordinated Zn^{2+} , whilst the RING-like take on a similar structure but without the Zn^{2+} co-ordination (Metzger et al., 2014). RING-type ligases mediate the attachment of ubiquitin via substrate recruitment and E2-Ub complex interaction, RINGs do not intrinsically transfer ubiquitin. RINGs can exist and function not only in a monomeric form, but also as a homodimer and heterodimer. It is interesting to note that each RING in homodimeric RINGs maintain individual activity with E2 whilst some heterodimeric RINGs do not. One RING will interact with E2 whilst the other serves to control activity, recruit substrates or stabilize the complex. One such example is the BRCA1- BARD1 complex, BCRA1 interacts with E2 and BARD1 lacks the crucial structural features necessary for E2 interactions (Brzovic et al., 2001). RINGs and U-box share common architectural features for E2 binding. Two loop regions surround a central α -helix groove providing a platform for E2 binding. RING-E2 binding consists of weak non covalent interactions through hydrophobic side chains. A key question in the

field was whether RINGs simply act as a mediator to E2-Ub and substrate, or whether they have the ability to enzymatically activate the E2-Ub complex. Initial crystallographic efforts revealed how RINGs position E2s relative to the substrates (Zheng et al., 2000) (Zhang et al., 2005) (Mace et al., 2008) (Yin et al., 2009) (Bentley et al., 2011). These structures did not reveal any conformational change to the E2, however the E2 was not charged due to the instability of the thioester bond. Later experimental efforts were designed to have E2 charged with ubiquitin through an ester or iso-peptide bond. Eventually, Plechanovová et al. and Dou et al. solved the complete structural complexes of dimeric RNF4 with UbcH5a-Ub and BIRC7 with UbcH5b-Ub, respectively. These structures revealed the E2-Ub complex in a closed conformation confirming RINGs activate E2s. This closed conformation geometrically favours the E2-Ub thioester bond for nucleophilic attack by the target substrate lysine (Pickart and Eddins, 2004). The structures showing RING ligase interactions with E2 and ubiquitin are shown in Figure 1.5. The mechanism for substrate recruitment is far more complex. Multiple RINGs can be active for the same substrate and multiple substrates can interact with the same RING ligase. Substrate control is conferred by a variety of processes such as domain recognition, post translational modifications and quaternary structure. Investigation into the role RING ligases have on cellular processes has uncovered their implications in human disease. The efforts to therapeutically target RING ligases are discussed in Section 1.2.3.2.

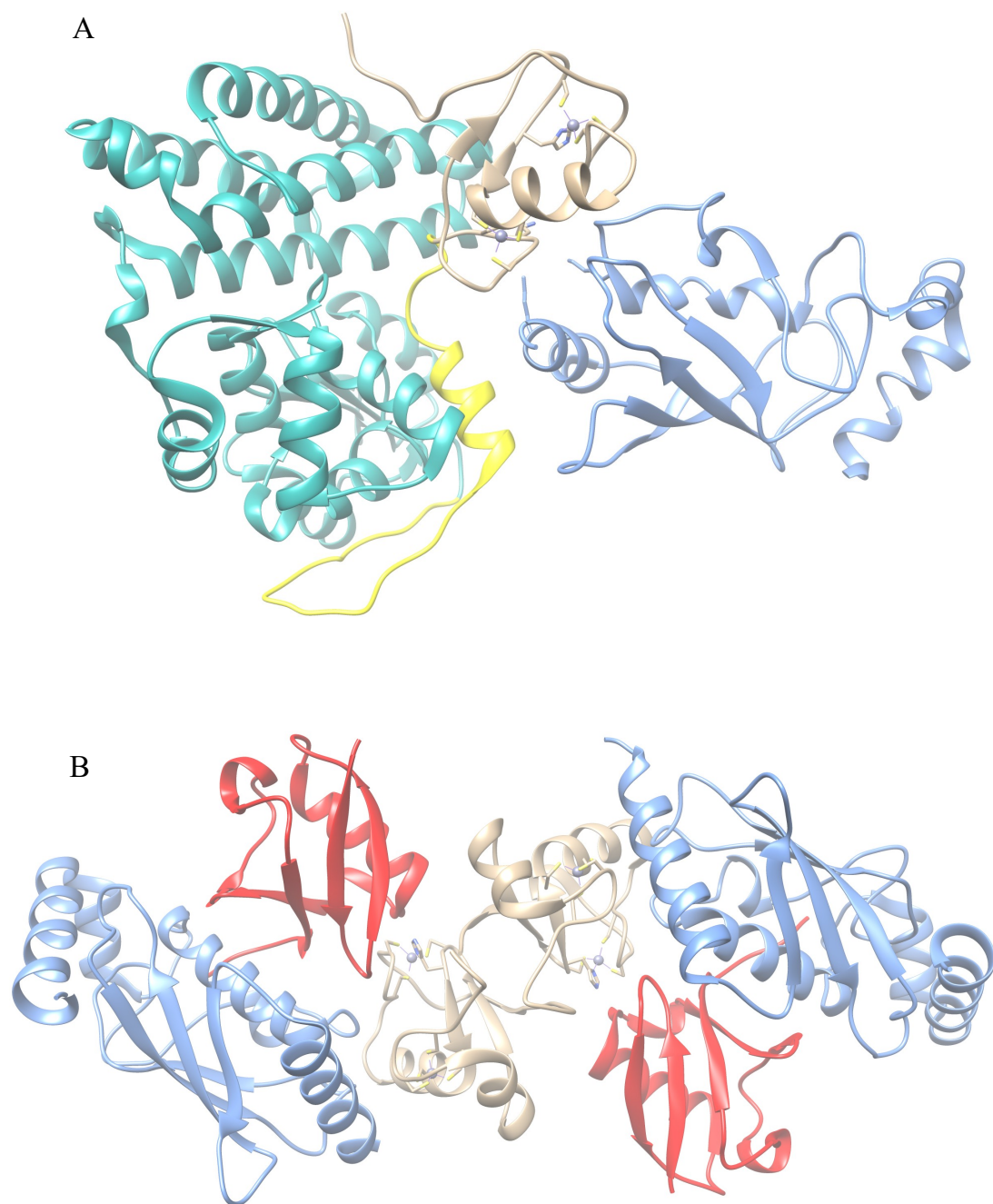


Fig. 1.5 -Structure of RING complexes (A) Structure of the interaction between the RING ligase c-Cbl and Ubch57 (blue). c-Cbl consists of a RING domain (brown), the substrate recognition domain TKB (turquoise) connected by the linker region (yellow). The structure was solved by x-ray diffraction to 2.9 Å (Zheng et al., 2000). PDB: 1FBV. (B) Structure of Rnf4 (brown)- Ubch5a (blue)- ubiquitin (red) heterotrimeric dimer. The structure was solved by X-ray diffraction to 2.21 Å. Image generated by UCSF Chimera (Pettersen et al., 2004).

1.1.3.2 HECT ligases

Unlike the RING ligases, HECT ligases intrinsically deliver ubiquitin onto target substrates by use of a catalytic cysteine. There are 28 known HECT E3 ligases, characterised by a C-terminal domain termed HECT (Lorenz, 2018). It is the HECT domain which holds the ability to form an E3:Ub complex during the final transfer of ubiquitin. The discovery of the E3 ligase E6AP led onto the first structural characterisation of the HECT domain (Huang et al., 1999). The HECT domain is bilobal, consisting of a large N-terminal lobe for E2 interaction and a smaller catalytic cysteine holding C-terminal lobe which are connected by a flexible linker region. The conformational flexibility between these lobes is critical to its ligase activity (Verdecia et al., 2003). Figure 1.6 shows a complex of E6AP and UbcH7. The N-lobe-E2 interaction is dominated by a key Phe residue in the loop region of E2, docking into a hydrophobic groove present in the HECT N-lobe. Ronchi et al. have presented evidence based on kinetic data for a second E2 binding site, however this has yet to be structurally defined. The flexibility of the hinge loop is clearly evident in the UbcH5b-Ub-Nedd4L complex where there is a change in E2-E3 conformation in order to bring the catalytic cysteines of E2 and E3 into close proximity, priming for the trans-thioesterification reaction (Kamadurai et al., 2009). Once the ubiquitin is charged onto the C-lobe, the ubiquitin C-lobe takes on a locked extended conformation, priming itself for nucleophilic attack from lysine residues of either target substrates, or further ubiquitin chains. Substrates can either be ubiquitinated in a processive or distributive manner. Processive refers to when substrates can be polyubiquitinated in one E3-substrate encounter, distributive refers to when only one ubiquitin can be attached during one enzyme- substrate encounter. Which mechanism of action dominates the HECT ligases is still subject to investigation, however Maspero et al. (2011) revealed a processivity site within Nedd4-1 that is used for building polyubiquitin chains. This site has later been targeted for enzyme inhibition by small molecules (Kathman et al., 2015). The domains located to the N-terminal of the HECT domain are responsible for

substrate recruitment. Based on the similarities and differences in these domains, the HECT group can be further be split into HERC, Nedd4 and the ‘other’ HECT E3s. The HERC family consists of 6 members containing the RLD domain. The subfamily is split into 2 large HERCs and 4 small HERCs. The two large HERCS contain two and three RLDs and the four small RLDs only have one RLD domain. HERCs have been implicated to be involved in a diverse range of cellular functions such as neurodevelopment and DNA repair (Sánchez-Tena et al., 2016). The ‘other’ HECT ligases do not share much commonality between each other and contain a wide range of N-terminal domains. The largest group is the Nedd4 subfamily, comprised of nine ligases. All nine ligases contain 2-4 WW domains, a C2 domain and the HECT domain. It is the Nedd4 ligases which form the basis of study for this thesis, and will be reviewed in greater detail.

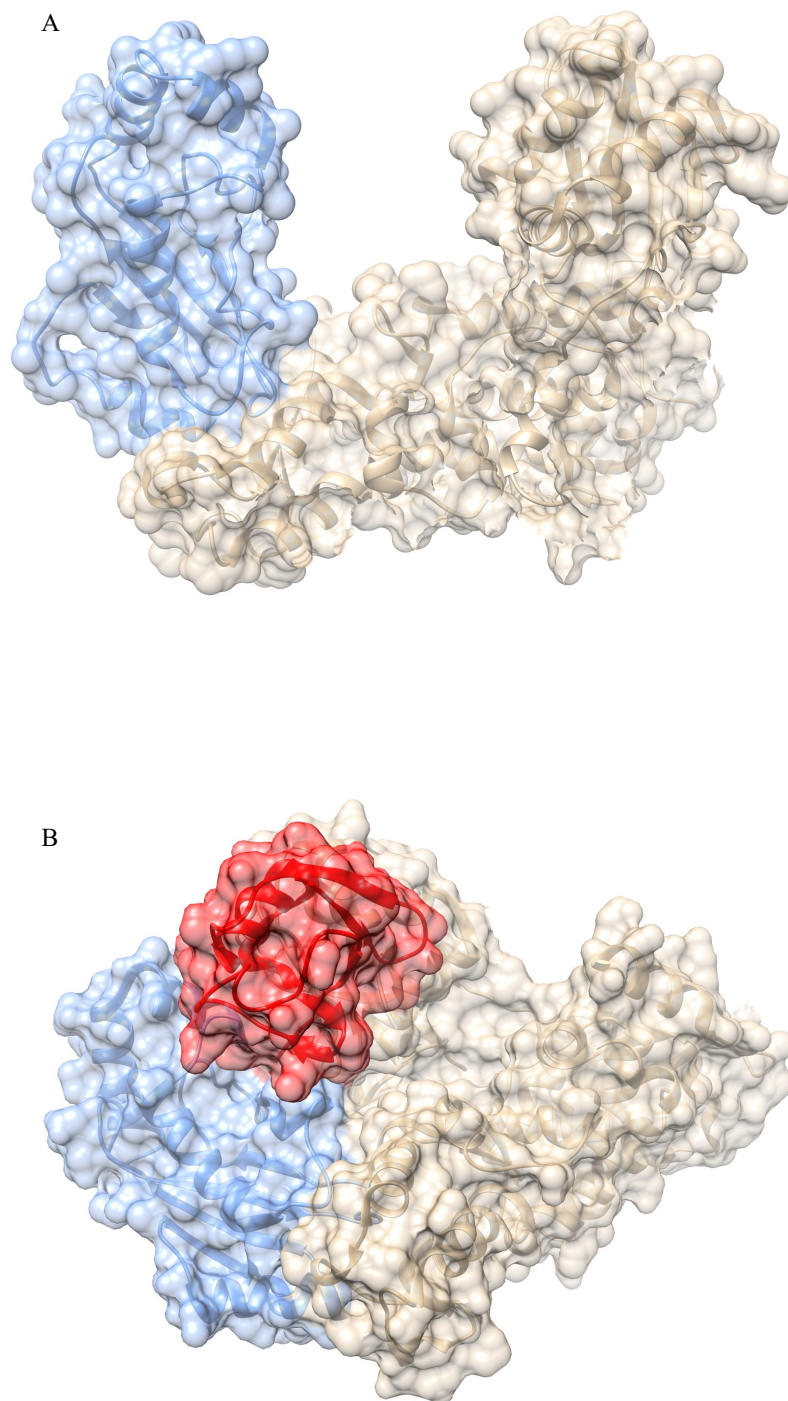


Fig. 1.6 -Structural representation of the L and T shape conformation of HECT(A) Bound structure of E6AP HECT domain (brown) and Ubch7 (blue). (B) Bound structure of Nedd4L HECT domain (brown) and Ubch5B (blue)- ubiquitin (red). Complexes solved by x-ray diffraction (Huang et al., 1999) (Kamadurai et al., 2009). PDB:1C4Z and 3JW0 respectively. Image generated by chimera (Pettersen et al., 2004).

1.1.3.3 Nedd4 E3 ligases

The Nedd4 supergroup of E3 ligases is made up of nine members: Nedd4 (also known as Nedd4-1), Nedd4L (Nedd4-2), Nedd4L1 (HECW1), Nedd4L2 (HECW2), WWP1, WWP2, Smurf1, Smurf2 and ITCH. The ligases share the common domain architecture of a N-terminal C2 domain, 2-4 WW domains and the HECT domain (Figure 1.7). The N-terminal C2 domain is an approximately 120 residue protein folded into 8 β -sheets in a β -sandwich. The C2 domain is responsible for membrane localisation and ligase/substrate trafficking by having the ability to interact with Ca^{2+} , phospholipids and membrane proteins (Rizo and Südhof, 1998). Plant et al. (1997) underlined the importance of the C2 domain as removal of the domain in Nedd4 resulted in loss of plasma membrane localization. The C2 domain is also a key regulator in controlling Nedd4 ligase activity, and solution NMR analysis showed the C2 domain of Smurf2 interacted with the catalytic c-lobe of HECT modulating its autoinhibition (Wiesner et al., 2007). Further structural studies elucidated these autoinhibition mechanisms to be that the C2 domain causes a conformational change of HECT into an inactive state. This inactive state prevents the Ub-thioester bond formation precluding substrate ubiquitination (Mari et al., 2014). The control in substrate recruitment and selectivity lies with the WW domains. They are termed as such as they contain two conserved tryptophan residues spaced 20-22 residues apart. The WW domain is made up of three antiparallel β -sheets forming a hydrophobic groove suited for ligand binding. WW domains interact with proline rich motifs in target substrates, more specifically Leu/Pro-Pro-X (any amino acid)-Tyr (L/PPXY) sequences (Chen and Sudol, 1995). It is increasingly evident the importance of the Nedd4 ligases in regulating a variety of aspects in cancer. Nedd4-1 is found to be overexpressed in human cancers such as breast, colorectal and lung cancer (Zou et al., 2015). Hansen et al. (2007) showed that the suppression of the ITCH gene in certain tumour cells can increase the effect of chemotherapeutic drugs. Smurf1 has shown to promote cancer metastasis in pancreatic, prostate, ovarian and breast cancer. The mis-

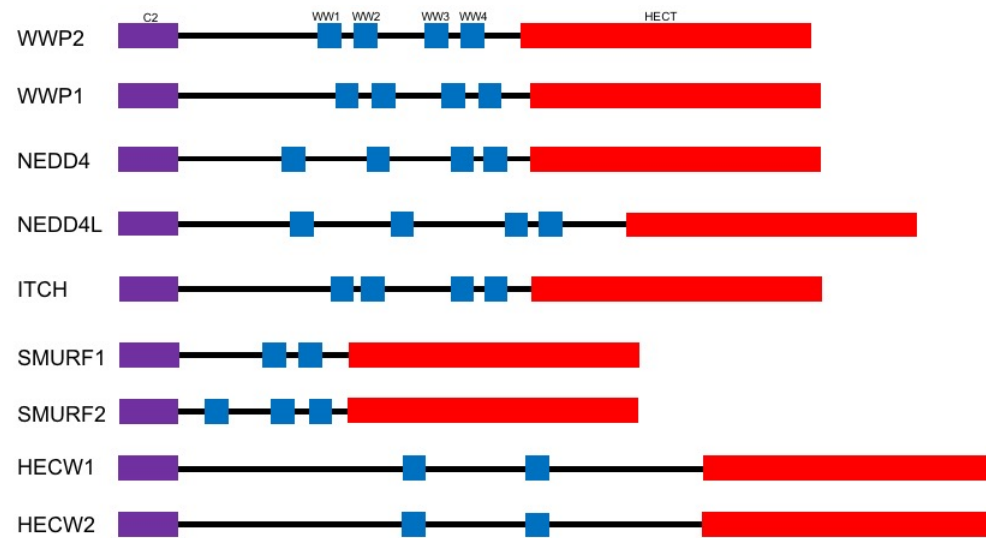


Fig. 1.7 -Schematic showing the Nedd4 superfamily of HECT E3 ligases. The C2 domain is represented in purple, the WW domains in blue and the HECT domain in red. Not to scale.

expression of Smurf2 has been found in breast, prostate, lymphoma, esophageal and renal cancers. WWP1 has frequently been found to be overexpressed in breast and prostate cancer (Bernassola et al., 2008). The work done by Soond et al. (2013) showed the mis-expression of WWP2 isoforms in breast, melanoma and prostate cancer cell lines.

1.1.3.4 WWP2: Structure, function and role in cancer

WWP2 was first identified in 1997 during a screen for WW domain containing proteins (Pirozzi et al., 1997). WWP2 is formed of the typical Nedd4 ligase architecture with a C2 domain, four tandem WW domains and the HECT domain. The HECT, WW2 and WW4 domains and the WW2-3 linker region has been structurally solved. The first WWP2 HECT structure was reported by Gong et al. (2015) (Figure 1.8) with the removal of five residues in the C-terminus necessary for producing high quality crystals. WWP2 HECT shares high homology with the HECT domain of WWP1, a structure-based sequence alignment revealed 83 % sequence similarity. The reported structure is missing residues 661- 702aa due to a lack of electron density, in which the authors predict is a result of intrinsic flexibility. The relative position of the N and C lobes is homologous to WWP1 HECT in adapting the inverted T conformation. During the discovery that the polypeptide chain between the WW2 and WW3 domains (termed 'linker region') acts as an autoinhibitory module; the WWP2 WW2 domain and 2,3 linker were structurally solved (Chen et al., 2017). WWP2 exists as three isoforms: full length (FL, all domains), N-terminal isoform (-N, WW1 and C2 domain) and C-terminal isoform (-C, WW4 and HECT).

All three isoforms have been linked in the growth and development of cancer and interestingly have differential activity in oncogenic pathways (Soond et al., 2013) WWP2 has been found to interact with PTEN and Smads, both key regulators in oncogenic signalling pathways. PTEN is a lipid phosphatase that acts as a tumour suppressor protein. PTEN acting as a tumour suppressor has been linked to have a key role in prostate cancer, with up to 70 % of primary prostate cancer tumours losing a copy of the PTEN gene. Maddika et al. found that the ubiquitin dependent degradation of PTEN is mediated by WWP2. The regulation of PTEN by WWP2 has been linked to tumour growth and metastasis. Cancer genomic analysis of patients (Figure 1.9) with prostate cancer shows almost 40 % of patients with castration resistant prostate cancer have the WWP2 gene amplified. A Kaplan-Meier plot of

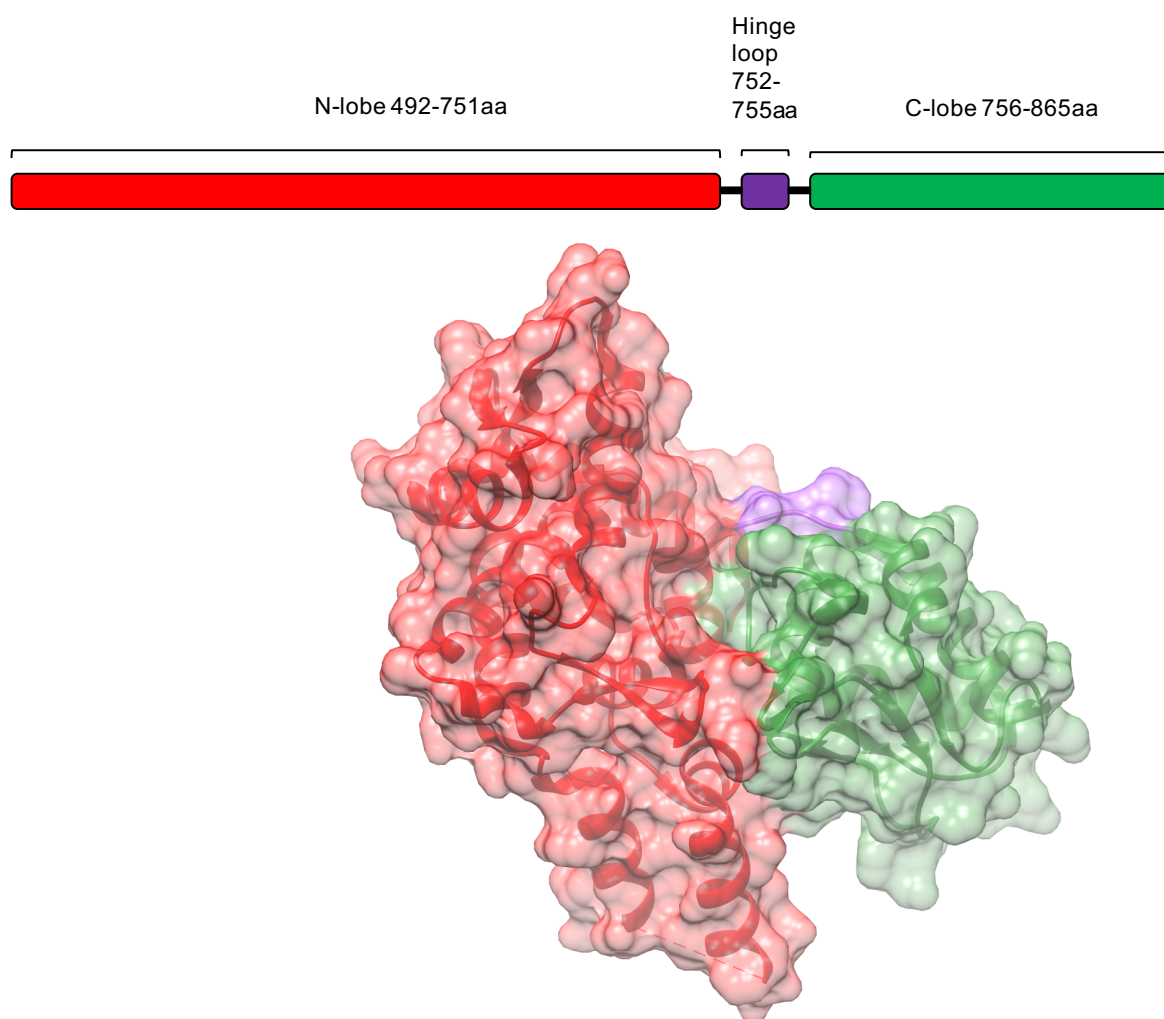


Fig. 1.8 -Structure of WWP2 HECT domain resolved to 2.51 Å (PDB: 4y07).

survival percentage shows there is a decreased chance of survival in patients over-expressing WWP1/2.

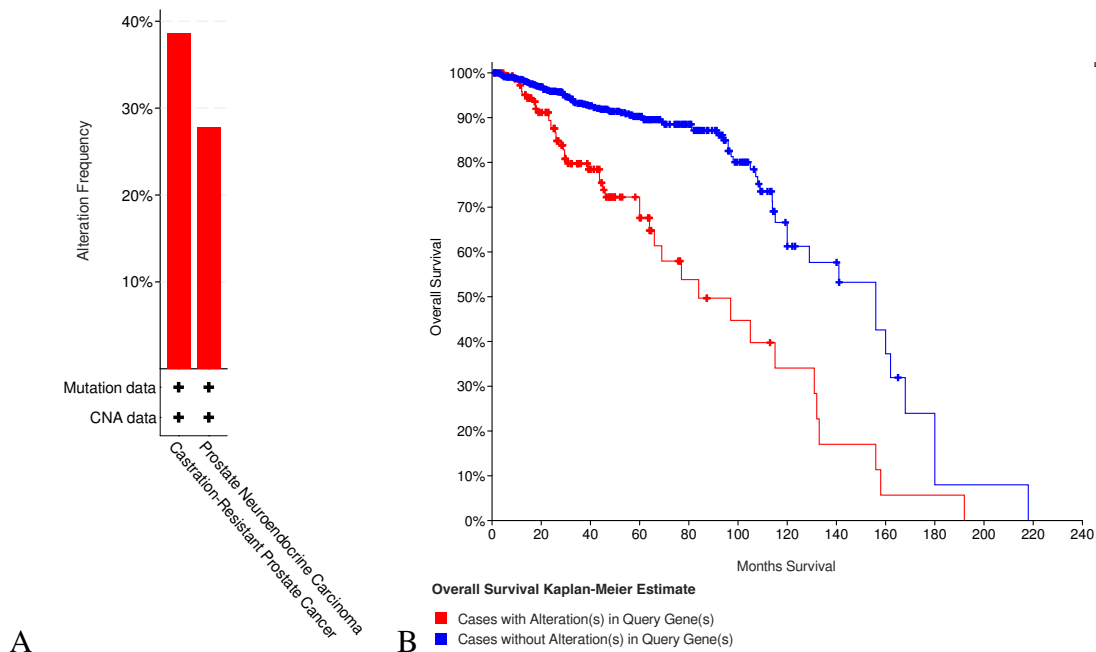


Fig. 1.9 -Genomic analysis of WWP2 and the highly homologous WWP1 in prostate cancer.(A) 38.57 % of 70 patients with castration resistant prostate cancer and 27.78 % of 54 patients with prostate neuroendocrine show amplification of WWP2/1. (B) Kaplan-Meier plot showing estimated survival. Blue shows patients without WWP1/2, red shows patients with WWP1/2. Data retrieved from a combined study of 4180 patients accessed through cBioPortal (Cerami et al., 2012) (Gao et al., 2013). Figure rendered by the cBioPortal platform.

1.2 Therapeutic targeting of the UPS

Over the past 25 years, there has been an increasing effort in targeting the ubiquitin proteasome system for therapeutic purposes. There is research focused on all aspects of the UPS regarding the use of small molecules as inhibitors. Since the first FDA approved proteasome inhibitor Bortezomib, research has extended to the enzyme cascade with E3 ligase inhibitors now in clinical trials. Recently, the pathway itself has been hijacked to ubiquitinate target proteins of choice- PROTACS. This section will review the development of small molecule inhibition of the UPS for cancer therapy.

1.2.1 Proteasome inhibitors

Those proteins targeted for degradation await the fate of the proteasome. The proteasome breaks down polyubiquitinated proteins into their constituent polypeptides. It is comprised of two subunits, the 19S particle directing the deubiquitination of substrates and the 20S core particle responsible for protein breakdown. The role the proteasome holds in cell proliferation and in cell apoptosis made it an interesting target for inhibition. Peptide aldehydes were the first class of proteasome inhibitors discovered (Vinitsky et al., 1992), followed by peptide vinyl sulfones (Bogyo et al., 1997) and the more potent peptide boronates (Adams et al., 1998). The boronate compound Bortezomib became the first FDA approved proteasome inhibitor for treatment of multiple myeloma in 2003. Natural products such as lactacystin (Fenteany et al., 1995), salinosporamide (Gulder and Moore, 2010) and epoxomicin (Meng et al., 1999) are known to inhibit proteasomal activity. The discovery of epoxomicin led to the derivative Carfilzomib being the second FDA approved proteasomal inhibitor. Lactacystin was the first non-peptide inhibitor of the proteasome and subsequent derivatives are the subject of further organic synthesis (Bulman Page et al., 2017) and ongoing clinical trials

(Teicher and Tomaszewski, 2015). The chemical structures of selected proteasome inhibitors are shown in Figure 1.10.

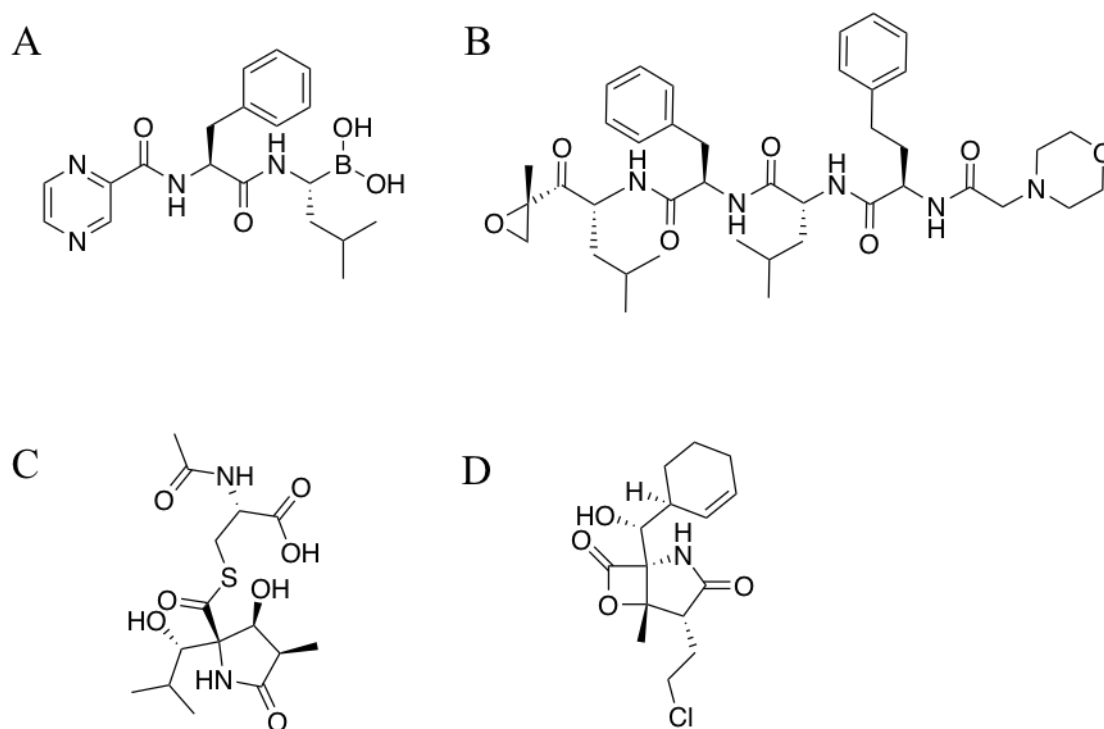


Fig. 1.10 -Selection of proteasome inhibitors (A) Bortezomib. FDA approved for multiple cell myeloma and mantle cell lymphoma. (B) Carfilzomib. FDA approved for multiple myeloma. (C) Salinosporamide A (also known as Marizomib). Currently undergoing clinical trials for multiple myeloma. (D) Lactacystin. The first non-peptide based proteasome inhibitor.

1.2.2 E1 and E2 inhibitors

During a screen for MDM2 dependent p53 ubiquitination, the compound PYR-41 was identified as an E1 inhibitor (Yang et al., 2007). PYR-41 is a pyrazolone based compound where the nitro group on the furan ring is suggested to block the catalytic cysteine of E1. The E1-Ub thioester formation has also found to be disrupted by the nitric oxide producing prodrug JS-K (Kitagaki et al., 2009). There remains concern for cysteine attacking inhibitors to interact with other thiol dependant enzymes. PYR-41 was tested against E2 and other catalytic cysteine depending enzymes and was found to have no activity. However PYR-

41 can partially reduce HECT E3 ligase activity, maintaining specificity concerns (Yang et al., 2007). E2 inhibition can be achieved either through interactions in the active site or allosterically. Ceccarelli et al. found that the compound ‘CC 0651’ (Figure 1.11) allosterically inhibited the E2 enzyme hCdc34 by binding in a site located 19 Å away from the active site. The binding perturbed the secondary structure of E2, resulting in the loss of ability to unload ubiquitin onto substrates. The inhibitor BAY 11-7082 on the other hand is found to covalently modify the catalytic cysteine of Ubc13 via a Michael addition with the vinyl sulfone (Strickson et al., 2013).

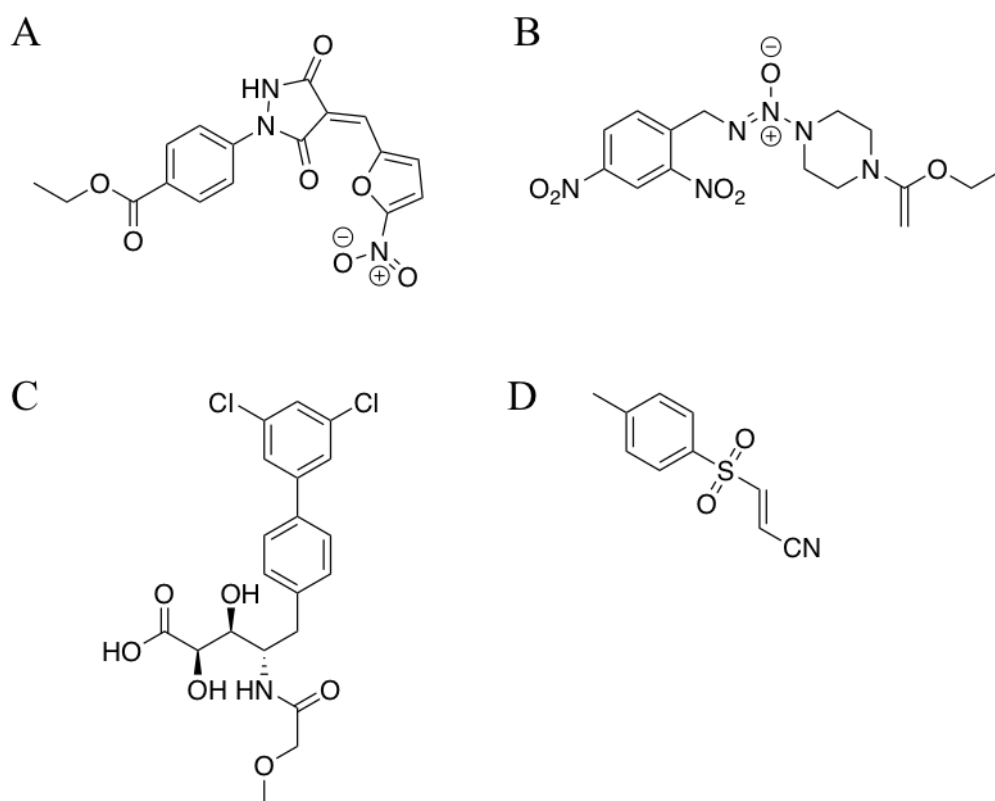


Fig. 1.11 -Small molecule inhibitors of E1 and E2. (A) The E1 inhibitor PYR-41 (B) The E1 inhibitor JS-K. (C) The E2 inhibitor CC0651 (D) The E2 inhibitor BAY 11-7082.

1.2.3 Targeting the E3 ligases

Targeting the final step in the cascade offers the greatest level of specificity and control. This approach is desirable in developing cancer treatments as inhibiting a single E3 ligase could potentially lead to less side effects. This section reviews the current research in targeting E3 ligases. However, one cannot discuss the current landscape of E3 ligase inhibitors without mentioning proteolysis targeting chimeras (PROTACs). This section will briefly describe PROTACs before going into more depth about the current E3 ligase inhibitors, as the overall scope of this thesis is driven towards 'traditional' small molecule inhibitor discovery.

1.2.3.1 PROTACs

There is an emerging interest in hijacking the ubiquitin proteasome pathway to target specific proteins. Instead of inhibiting a target proteins activity with a small molecule, the UPS is used to remove the target protein (Sakamoto et al., 2001). This is done by using a hetero-bifunctional molecule consisting of two ligands joined by a linker group; the PROTAC. One ligand is specific for the target protein and the other is specific for the E3 ligase of choice. The PROTAC can simultaneously bind to the E3 ligase and target protein, forming a ternary complex. This recruiting of an E3 ligase causes ubiquitin from E2 to be transferred onto the target protein, consequently labelling it for its degradation. The main appeal to this strategy is that 'undruggable' proteins can be targeted, such as proteins without enzymatic activity (Crews, 2010). All of the E3 ligases used in PROTACS are currently from the RING group of ligase. The main ligases used are MDM2, cIAP1, CRBN and VHL (Burslem and Crews, 2017). There is a need to discover more ligands for E3 ligases owing to the limited number of E3 ligands available. Whilst the focus of these PROTACs have mainly been for oncology, there is a desire to move to longer term chronic diseases. The limited number of ligands makes this difficult owing to the undesirable properties of the ligands (e.g. the CRBN inhibitor thalidomide being teratogenic). It would be interesting to investigate whether

the HECT ligases still have the ability to ubiquitinate substrates via a PROTAC. A recent comprehensive review of current PROTACs can be found at Wang et al. (2019).

1.2.3.2 RING inhibitors

RING ligases have been the more extensively studied ligase for small molecule inhibition. The RING ligase MDM2 is known to target the tumour suppressor p53. Nutlins were the first class of compounds identified as MDM2 inhibitors and are now undergoing early stage clinical trials (Burgess et al., 2016). They are a collection of cis substituted imidazole analogues. Nutlins bind to the p53 binding site of MDM2, disrupting the protein- protein interactions. The VHL complex is another RING ligase in which extensive medicinal chemistry is currently ongoing (structural development shown in Figure 1.12). Through the process of structure based drug design, inhibitors with initial moderate potency have been developed to have nanomolar activity in cells (Buckley et al., 2012) (Galdeano et al., 2014) (Soares et al., 2017). Although there are still few RING ligases which have been targeted, inhibitor development is further through the pipeline than the HECT ligases.

1.2.3.3 HECT inhibitors

The first small molecule inhibitor for HECT was reported in 2014 by Mund et al. (2014). Originally, a phage display approach was used to identify bicyclic peptides as specific inhibitors for Smurf2, Nedd4, Huwe1 and WWP1. However, work with these peptides could not progress beyond *in vitro* as they could not be delivered into cells. Intuitively, the authors exploited the binding properties of the peptides and used Alpha Screen technology to search for small molecules which would disrupt the HECT-peptide interaction. After a screen of around 17500 compounds, the molecule termed 'heclin' was identified as the leading inhibitor, with an IC₅₀ of 6-7 μ M for several HECT ligases. Hydrogen-deuterium exchange and mutational studies concluded a predicted model of action where heclin binds

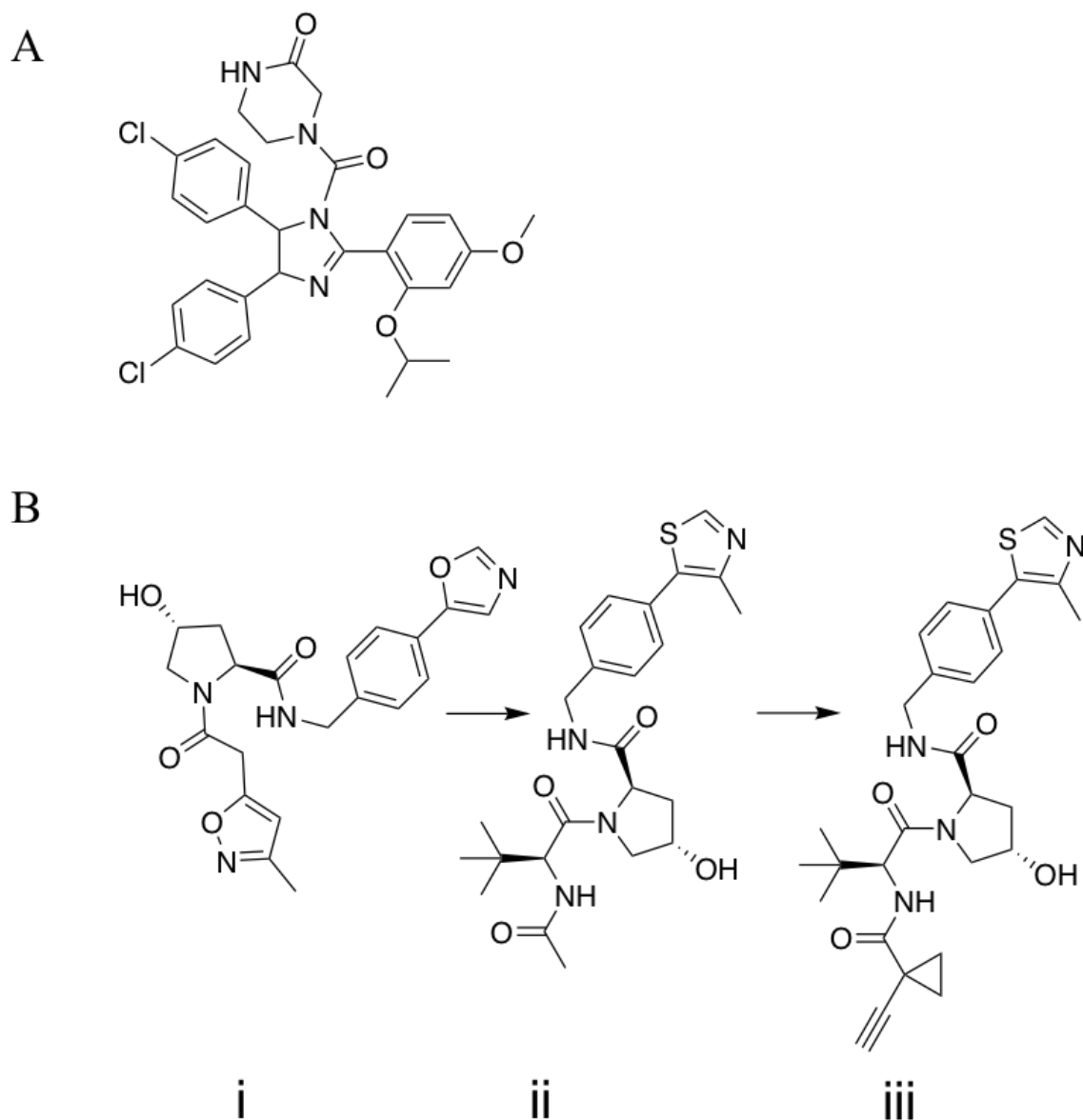


Fig. 1.12 -Inhibitors of E3 RING ligases. (A) Nutlin 3, the MDM2 inhibitor currently in clinical trials. (B) (i) The first small molecule targeting the VHL. (ii) Using a structure- based design approach, this compound exhibited nanomolar affinities. (iii) Optimised structure showed improved cellular permeability.

to HECT causing a conformational change within the C-lobe to expose the catalytic cysteine. Exposure of the catalytic cysteine makes it prone to oxidation, rendering its ligase activity defunct. The location of the heclin binding site is not known, but the cleft between the C-lobe and E2 binding site is stated as a possible contender. Despite showing selectivity

for HECT ligases over RING ligases, there is no specificity between each individual HECT ligase. Furthermore, the heclin molecule contains medicinal chemistry red flags, such as the metabolically unstable furan ring and the acrylamide Michael acceptor (Chen et al., 2018). The potential promiscuity and metabolic stability *in vivo* remains to be investigated. It is interesting to note that in a recent study reporting WWP1 as a biomarker for acute myeloid leukemia, heclin was found to suppress leukemic blast growth (Sanarico et al., 2018). This exemplifies the importance for research into HECT ligase inhibitors as therapeutics. Rossi et al. (2014) used a high throughput ELISA based assay to screen for inhibitors of ITCH autoubiquitination and found the antidepressant drug clomipramine to inhibit ITCH autoubiquitination and ITCH mediated p73 ubiquitination. Clomipramine analogues were tested in the same ELISA assay and small structural alterations had a significant effect. Increasing the alkyl chain on the terminal side chain amine decreased levels of inhibition. In contrast the inhibitory effects could be increased fivefold by addition of a chloride on the phenyl ring. This is in accordance with the proposed mechanism of action which is driven by a nucleophilic aromatic substitution reaction with the catalytic cysteine and chloro substituted phenyl moiety of clomipramine. Structures of clomipramine and heclin are shown in Figure 1.13.

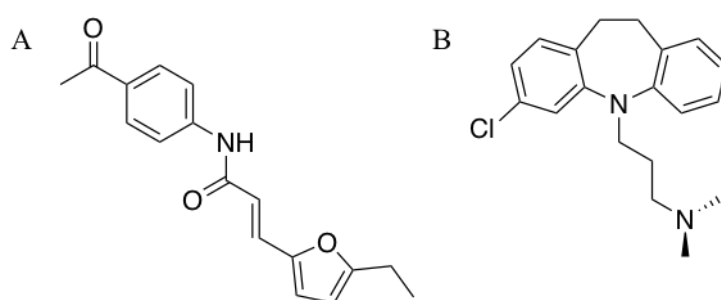
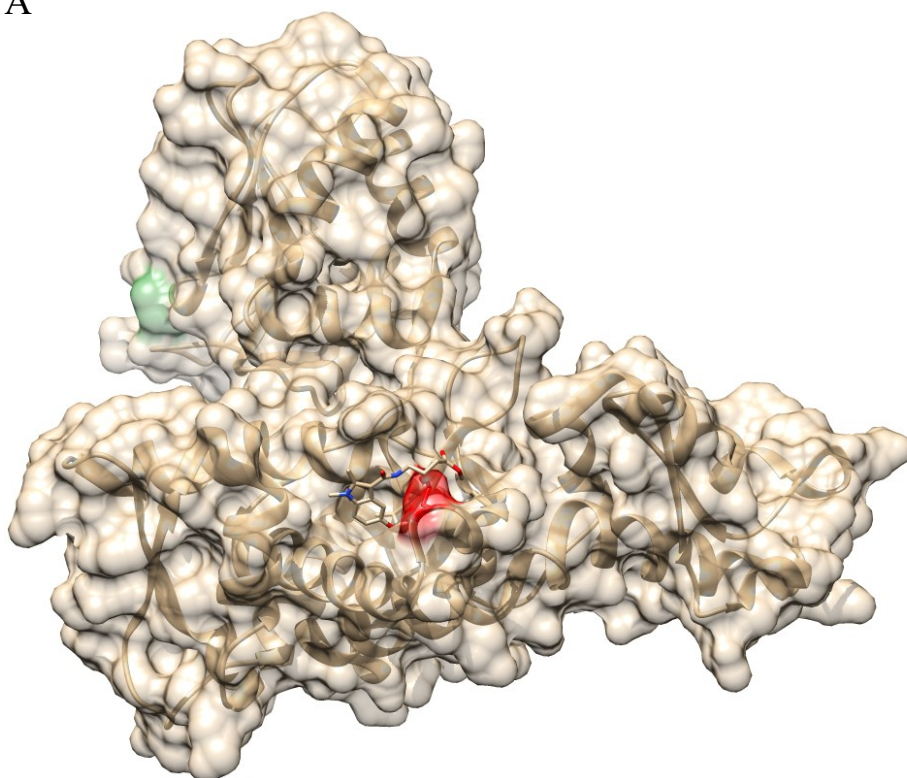


Fig. 1.13 Inhibitors of E3 HECT ligases (A) Heclin, the first HECT inhibitor. (B) Clomipramine, the antidepressant inhibitor of ITCH.

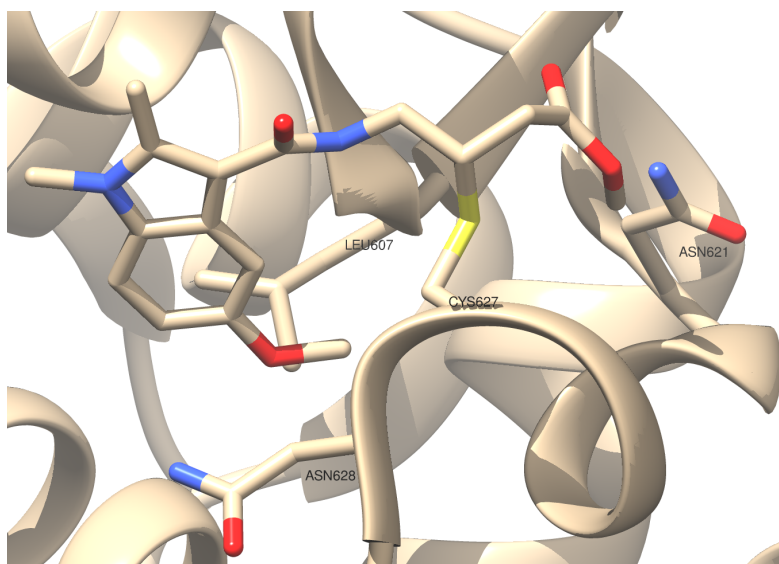
The first crystal structure of a bound HECT ligase inhibitor complex came in 2015 from Kathman et al. (2015) (Figure 1.14). The group used their previously reported ‘covalent

tethering' approach which screens electrophilic fragments against cysteine residues (Kathman et al., 2014). These electrophilic compounds act as Michael acceptors and covalently bond with the thiol group present in cysteines. Nedd4 was treated with a mixture of electrophilic compounds, identifying 2 compounds (Figure 1.14) as direct hits to the non catalytic Cys627. A crystal structure was obtained with Nedd4 HECT and "compound 1" (Figure 1.12), elucidating the binding mode of this compound. The structure clearly shows the formation of a covalent bond between the ligand and Cys627. The α , β - unsaturated carbonyl sidechain acts as the Michael acceptor to the reactive thiol. The aromatic edge to face interactions between the indole moiety and the local tyrosine provide stability to the complex. As this compound sits in the allosteric steric site involved in non covalent bonding with ubiquitin during the build up of polyubiquitin chains, it was theorized this compound would disrupt Nedd4 HECT: Ub interactions and inhibit ubiquitination. However, when incubated with the approximate concentration of ubiquitin found in cells (approximately 60 μ M) the ligand was inactive. In the search for more potent analogues it was found substitutions were tolerated on the *N*-methyl of the indole, so a series of *N*- substituted indoles were synthesised. The cyclopentane derivative was the most potent in disrupting the Nedd4 HECT: Ub interaction with a K_I of 29.3 μ M.

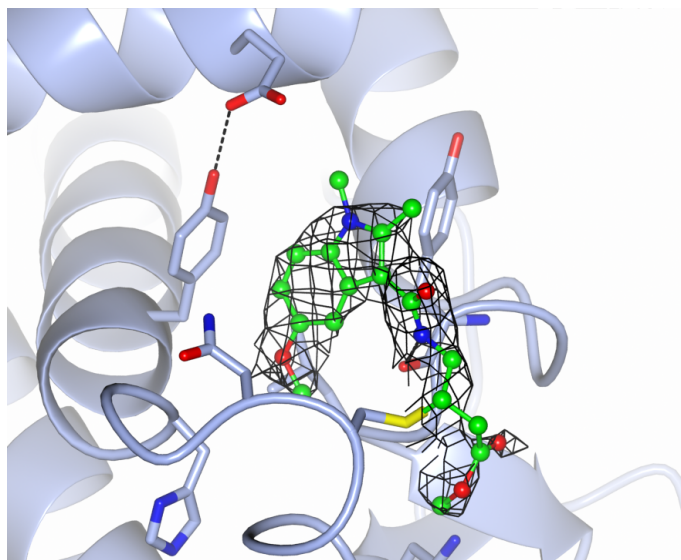
A



B



C



D

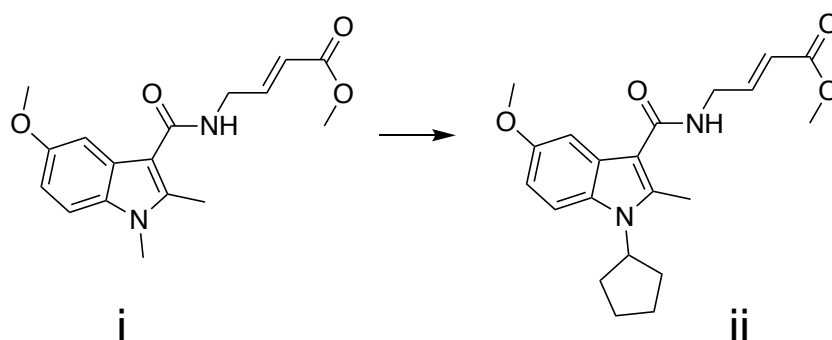


Fig. 1.14 -Structure of the Nedd4 inhibitor. (A) Structure of compound 1 bound to Nedd4. Compound 1 is covalently bound to N lobe Cys657, shown in red. The catalytic Cys867 is shown in green. (B) Zoomed in representation of the binding pocket. (C) Electron density (black mesh, contoured at 1σ) of compound 1 bound to Cys627. Image generated using CCP4MG (McNicholas et al., 2011). (D) (i) The initial *N*-methylated compound, used in the crystallography experiments. Optimisation to ii by the substitution of the *N*-methyl to *N*-cyclopentane.

1.2.3.4 Other HECT ligase inhibitors

The natural product indole 3-carbinol has been showed to downregulate Nedd4 mediated PTEN ubiquitination (Aronchik et al., 2014). Quirit et al. investigated whether this compound was acting as a direct inhibitor on Nedd4. After finding the compound was interacting with the HECT domain via isothermal calorimetry, *in silico* docking using the previous Nedd4:inhibitor was performed to map the possible binding site. Using this structural information, a series of analogues was synthesised and tested for binding by protein thermal shift. The most potent analogue had an ELISA measured autoubiquitination IC₅₀ of 2.71 μ M. The indole carbinol compounds were also found to inhibit human melanoma cells to a level which correlated with their Nedd4 inhibitory activity. Smurf1 has been subjected virtual screening and cell assay studies in order to find novel inhibitors (Cao et al., 2014). Whilst these inhibitors could be interpreted to modulate Smurf1 activity, there is yet to be any structural studies. Moreover, some of these identified compounds also have the ability to covalently modify cysteines, meaning their specificity would have to be fully assessed. Zhang et al. (2016) utilised a phage display to search for ubiquitin variants that would have high affinity to certain HECT domains. A phage library encoding up to a billion ubiquitin variants which searched for high affinity between 20 HECT domains. Ubiquitins with both inhibitory and activator effects were identified with their target HECT sites structurally identified. Whilst not likely to have direct therapeutic potential, these UbV's can be used as biological tools to probe HECT mechanisms.

1.3 Thesis aims

The current state of drugging the ubiquitin proteasome pathway has been explored in this chapter. This landscape is rapidly growing with already FDA approved treatments and numerous compounds in all stages of the drug discovery pipeline. Targeting the E3 ligases not only gives the opportunity for traditional small molecule inhibitors, but to entirely new concepts such as PROTACs. The link between WWP2 and cancer has been accumulating evidence over the past few years, presenting itself as a promising novel cancer target. The previously discovered HECT inhibitors in the high homology Nedd4 family sets a precedent that WWP2 could be targetable. The presence of the enzymatic HECT domain with a catalytic active site and probable allosteric sites offers an appealing target for drug design. There are no inhibitors which target WWP2 and moreover there is no structural evidence for a non-covalent inhibitor for any of the Nedd4 ligases. The overall aim of this thesis is to design, synthesise and biologically evaluate novel inhibitors for the WWP2 enzyme. The design of the compounds will be through small molecule screening approaches collected through biophysical assays and other data gathered in the Chantry lab. The structural basis of inhibition will then be revealed through the solving of the first WWP2: small molecule X ray crystallographic structure. Not only will a crystallographic structure be used, but data from the newly invented DEEP STD will be exploited to generate an accurate picture of a binding site. This data will be collated to inform organic synthesis of a selected compound with the intention of improving activity. The synthesised compounds will then be biologically tested *in vitro* against WWP2 and Nedd4. This work hopes to progress the research into WWP2 inhibitors with the intention of using these discoveries to develop novel therapeutics targeting WWP2, which there are currently none known.

Chapter 2

Biological materials and methods

2.1 DNA techniques

2.1.1 Plasmids and primers

Table 2.1 describes the plasmids used during this project with the description and origin.

Plasmid	Origin
pGEX2T-WWP2-FL	Dr Andrew Chantry's lab
pOPINF-WWP2-FL	Dr Andrew Chantry's lab
pGEX-5X3-Nedd4-FL	Gifted by Shiaw-Yih Lin (Addgene plasmid #45043)
pET3a-UbcH7	Gifted by Professor Martin Scheffner
pET3a-hUBa1	Gifted by Titia Sixma (Addgene plasmid #63571)
pGEX6p-2-WWP2-WW2-2,3-linker-HECT	Gifted by Professor Philip Cole
pET28a-WWP2-HECT	Cloned during this work by In-Fusion
pET28a-WWP1-HECT	Cloned during this work by In-Fusion
pET28a-Nedd4-HECT	Cloned during this work by In-Fusion

Table 2.1 -Plasmids used throughout this project

Primers (obtained from Eurofins Genomics) designed for In-Fusion cloning contain 15bp preceding the restriction enzyme site from the pET28a vector were added to the beginning of the gene specific sequence. Primers were designed for ligation between the restriction sites Nde1 (sequence CATATG) and Xho1 (sequence CTCGAG) in the pET28a vector. Primers were re-suspended from their lyophilized form in nuclease free d_4H_2O to obtain a concentration of 100 nmol mL⁻¹. They were further diluted to a concentration of 10 nmol mL⁻¹ for PCR. The primers used for this work are described in Table 2.2.

PCR product	Primers 5'-3'	Plasmid
WWP1 HECT	FWD- GTGCCGCGCGGCAGCCATATGATGGGCTTTAGGTGGAAGCTTG REV- GTGGTGGTGGTGGTGCTCGAGTTACTCTGTCTCTTCTATTGCAAAAAG	pET28a
WWP2 HECT	FWD- GTGCCGCGCGGCAGCCATATGATGGACCGCAGTTTTTCGG REV- GTGGTGGTGGTGGTGCTCGAGTTACTCGGTCTCCTCAATGG	pET28a
Nedd4 HECT	FWD- GTGCCGCGCGGCAGCCATATGATGTCCAGGGATTACAAAAGAAAG REV- GTGGTGGTGGTGGTGCTCGAGTTAATCAACTCCATCAAAGCCCTG	pET28a

Table 2.2 -Primer sequences

2.1.2 Phusion PCR

Phusion High- Fidelity DNA polymerase reactions were set up as described in Table 2.3. The cycling conditions used are described in Table 2.4 . Amplification was confirmed by running the DNA samples on 1 % TAE- agarose gel stained with SYBER Safe (Invitrogen) at a final concentration of 0.01 %. For larger scale PCR reactions, 5 x 20 μ L reactions were carried out. The resulting solution was mixed and run on agarose gel. The DNA was gel extracted and purified.

Component	20 μ L pcr reaction
dH ₂ O	12.2
HF Buffer	4
10 mM dNTPs	0.4
Forward Primer	1
Reverse Primer	1
Template DNA	0.2
DMSO	1
DNA Phusion polymerase	0.2

Table 2.3 -PCR reaction mixture

Cycle	Temperature (°C)	Time (s)	Number of cycles
Initial denaturation	98	120	1
Denaturaton	98	10	30
Annealing	63	30	30
Extension	72	45	30
Final extension	72	600	1

Table 2.4 -PCR cycle conditions

2.1.3 DNA Gel electrophoresis

DNA samples were mixed with 6 x DNA loading dye (Thermo Scientific) at a final concentration of 20 %. Samples loaded onto the gel were run at 100V for 60 minutes.

2.1.4 Plasmid digest

The pET28a vector was digested between the restriction sites NdeI and XhoI using the conditions described in Table 2.5.

Component	NdeI digest (μ L)	XhoI digest (μ L)	XhoI + NdeI digest (μ L)
H ₂ O	8.3	8.3	14.6
Restriction enzyme	1	1	2+2
BSA	0.2	0.2	2
pET28a DNA	8.5	8.5	0.4
Buffer D	2	2	4

Table 2.5 -Restriction enzyme digest reaction

The reaction mixture was left to digest for 4 hours at 37 °C. Agarose gel electrophoresis was run to confirm the digested vector.

2.1.5 In-Fusion ligation

In-Fusion[®] HD cloning was used for the ligation of the PCR product into pET28a. The reaction was set up as described in Table 2.5 and carried out according to the manufacturers protocol. DNA was transformed into Stellar competent cells and 3- 6 colonies picked for plasmid extraction.

Component	Amount (μ L)
5 x In-Fusion [®] HD Enzyme Premix	2
Digested pET28a	3
Insert DNA	1
dH ₂ O	4

Table 2.6 -Plasmid ligation reaction mixture

2.1.6 DNA extraction and purification

DNA was cut from the agarose gel using a scalpel and purified using the QIAquick Gel Extraction kit (Qiagen) according to manufacturers protocol.

2.1.7 DNA transformations

The appropriate cells were thawed on ice for 10 minutes. 1 μ L of desired DNA was mixed into 50 μ L cells and incubated on ice for 30 minutes. Cells underwent heat shock for 45 seconds at 42 °C and placed immediately on ice for 10 minutes. 150 μ L LB was added and incubated at 37 °C for 45- 60 minutes. 150 μ L of the mixture was plated onto the appropriate antibiotic LB agar plate. The plate was incubated overnight at 37 °C.

2.1.8 Plasmid purification

Plasmids were purified using either the QIAprep Spin Miniprep kit or the QIAprep Spin Midiprep kit depending on scale. The concentration of DNA was measured using a NanoDrop spectrophotometer (Thermo Scientific).

2.2 Protein techniques

2.2.1 Protein expression

A single colony from the transformation plate with the desired plasmid was picked and inoculated with the desired antibiotic in LB. The starter culture was incubated overnight at 37 °C. Starter culture was added to a larger volume of LB in a ratio of 1:25 to 1:50 and grown to OD₆₀₀ 0.4-1.0. IPTG was added and incubated in the optimum conditions for each construct as stated in Table 2.7. Cells were pelleted and stored at -80 °C.

Protein	Cell line	IPTG concentration	Induction temperature	Induction length
pGEX2T- WWP2 FL	CodonPlus RP	0.75 mM	30 °C	O/N
PopinF- WWP2 FL	CodonPlus RP	1.0 mM	16 °C	O/N
pET28a- WWP2 HECT	Rosetta	1.0 mM	15 °C	O/N
pGEX6p-2- WWP2 WW2-2,3 linker- HECT	CodonPlus RP	0.5 mM	16 °C	O/N
pET28a- WWP1 HECT	CodonPlus RP	1.0 mM	30 °C	6 hours
pGEX 5X3- Nedd4 FL	Rosetta (DE3)	1.0 mM	30 °C	O/N
pET28a- Nedd4 HECT	BL21 Star (DE3)	1.0 mM	30 °C	O/N
pET3a-UBA1	BL21(DE3)	1.0 mM	25 °C	O/N
pET3a- UBH7	CodonPlus RP	1.0 mM	25 °C	4 hours

Table 2.7 -Protein expression conditions

2.2.2 Protein purification

2.2.2.1 His tagged proteins

Pelleted cells were resuspended in 35 mL lysis buffer and sonicated until sufficient cell lysis had been reached. The sample was centrifuged at 18 000 rpm for 45 minutes at 4 °C using the Beckman Ja20 rotor. The clarified lysate was loaded on a prepacked 5mL Ni sepharose HisTrap HP using a peristaltic pump (Parmachia Biotech). His tagged proteins were loaded onto an AKTA Pure system. Bound protein was eluted with a gradient of Ni NTA elution buffer. Eluted samples were analysed by SDS-PAGE gel electrophoresis and samples containing the desired protein were pooled. Buffer recipes are shown in table 2.8.

2.2.2.2 GST tagged protein

Clarified lysate was obtained in the same manner as His tagged proteins. Buffers are described in Table 2.9. Clarified lysate was loaded onto a GSTrap FF (GE Healthcare Life Sciences) using a peristaltic pump. The column was washed with 25 mM Tris Cl pH 8.0, 250 mM NaCl, 0.1 Triton X-100 before being eluted in 30 mL GST elution buffer.

2.2.2.3 Dialysis

Protein solution was dialysed using 10 kDa molecular weight cut off SnakeSkin (Thermo Fisher Scientific) in a 5 L reservoir of the desired buffer. Fresh reservoir solutions were used at least twice.

2.2.2.4 GST cleavage

PreScission Protease (Thermo Fisher Scientific) was added to a concentration of 1U per 1mg of GST tagged protein in a buffer consisting of 25 mM Tris-Cl pH 8.0, 250 mM NaCl and 5 mM DTT and incubated overnight at 4 °C. Uncleaved protein and GST was removed by running the sample down a GS column.

2.2.2.5 Gel filtration

The protein sample was concentrated to 2 mL using a 3 KDa MW cut off Vivaspin protein concentrator (GE Healthcare). The 2 mL sample was loaded onto the Sephadex 16/600 column through a 2 mL sample loop. A standard run of 1 mL min⁻¹ for two column volumes was set up. The protein was eluted in the buffer described in Table 2.9. The eluted samples were monitored by UV absorbance at 280 nm. Analysis by SDS-PAGE on the peak fractions confirmed the correct molecular weight of the target protein.

Ni NTA lysis buffer	Ni NTA elution buffer
50 mM Tris- Cl pH 8.0	50 mM Tris- Cl pH 8.0
500 mM NaCl	500 mM NaCl
10 mM imidazole	250 mM imidazole
10 % Glycerol (v/v)	10 % Glycerol (v/v)
5 mM DTT	5 mM DTT

Table 2.8 -Ni NTA purification buffer recipes

GST lysis buffer	GST elution buffer	Gel filtration buffer
25 mM Tris-Cl pH 8.0	25 mM Tris-Cl pH 8.0	25 mM Tris-Cl pH 7.5
250 mM NaCl	250 mM NaCl	150 mM NaCl
1 mM PMSF	50 mM Reduced glutathione pH 8.0	5 mM DTT
1x Roche protease inhibitor cocktail		

Table 2.9 -GST purification buffer recipes

2.2.3 Protein crystallisation

2.2.3.1 Crystal formation

Crystallisation trials were set up by using the sitting drop vapour diffusion method in MRC 48 well plates or MRC 96 2 drop plates. For the 48 well plates, 1 μL of protein samples at a concentration of 1-10 mg mL^{-1} was mixed with 1 μL of reservoir solution. The adjacent well contained 200 μL of reservoir solution. For the 96 well plates, 0.5 μL of protein was mixed with 0.5 μL reservoir solution and the adjacent buffer well contained 100 μL of reservoir solution. The plate was sealed with optically clear sheets (Molecular Dimensions or 4titude) and crystals were left to form at 4 $^{\circ}\text{C}$ or 16 $^{\circ}\text{C}$. Crystal growth was monitored using a SZX9 Stereo Microscope (Olympus).

2.2.3.2 Crystal harvesting

Single crystals were picked using round LithoLoops at 16 $^{\circ}\text{C}$. Crystals were soaked in selected ligands and/or cryoprotected as outlined in Chapter 4 before being subjected to liquid nitrogen for storage.

2.2.3.3 X-ray data collection

Protein crystals were sent to the synchrotron at Diamond Light Source for data collection. Beamlines IO3 and IO4 were remotely controlled for data collection. Collected data was

processed using the available diamond pipelines and downloaded for subsequent molecular replacement and refinement using PHENIX and COOT respectively.

2.2.4 SDS-PAGE gel electrophoresis

Samples for SDS-PAGE were prepared by adding laemmli (eq. volume) and DTT (10 mM final concentration). To denature the proteins, the samples were heated to 98 °C for 3 minutes. Samples were analysed on 10 % or 15 % acrylamide gels using the Mini-PROTEAN tank with a 1 % SDS running buffer for 60 minutes at 180 V. InstantBlue (Expedeon) was used to stain the gels and protein bands appeared after 15 minutes. Acrylamide gels were prepared according to the amounts described in Table 2.10.

Component	Resolving gel		Stacking gel 6 %
	10 %	15 %	
dH ₂ O	15.3 mL	11.3 mL	11.6 mL
37:5:1 40 % Acrylamide	8 mL	12 mL	
1.5 M Tris pH 8.8	8 mL	8 mL	
0.5 M Tris pH 6.8			5 mL
10 % SDS	320 µL	320 µL	200 µL
10 % APS	320 µL	320 µL	200 µL
TEMED	32 µL	32 µL	20 µL

Table 2.10 -10 % and 15 % acrylamide gel components and volumes for 2 gels

2.2.5 Plate coating

96 or 384 well plates (Corning™ Costar™ EIA) were coated with a 3 % bovine haemaglobin solution containing 0.05 M Na₂CO₃ and incubated at 4 °C overnight. The wells were washed 6 times with phosphate buffer solution (PBS) and the heterobifunctional cross-linker sulphosuccinimidyl 4-(p-maleimidophenyl)butyrate (SSMPB) was added to a final concentration of 0.1 mM in PBS per well. The plates were incubated for 2 hours at room temperature and washed with PBS 4 times. A solution of 150 mM NaCl, reduced glutathione 10 mM and 1 mM EDTA pH 6.7 was added to each well and incubated overnight at room temperature. The wells were washed with PBS containing 0.01 % Tween 20 (PBST). Coated plates were either used immediately, or stored in PBST at 4 °C. At least 50 µL was used for all components when using 96 well plates and at least 20 µL for all components was used when using 384 well plates.

2.2.6 Autoubiquitination assay

WWP2 Fl pelleted cells were resuspended in 10 mL PBS per 50 mL of expressed culture and sonicated until sufficient cell lysis. The cell lysate was centrifuged for 30 minutes at 4300 rcf (Heraeus Megafuge 1.0R). 100 µL of clarified lysate was added to each well of a glutathione coated 96 well plate and incubated for 1 hour at room temperature. E1 (20 ng per well), E2 (240 ng per well), ubiquitin (100 ng per well) and ATP (1.25 mM) in a buffer of 25 mM Tris pH 8, 100 mM NaCl, 4 mM MgCl₂ and 50 µM DTT were incubated at room temperature for 40 minutes. The plate was washed three times with PBST and tapped dry. 2 µL of the compound in question was added to the plate at the desired concentration in 10 % DMSO. 18 µL of the ubiquitination mix was added to each well and the plate was incubated at room temperature for 2 hours. The plate was washed with PBST. 100 µL of anti-FLAG HRP (Sigma) (1:10000 PBST) was added to each well and incubated for a further hour at

room temperature. The plate was washed. 100 μ L of TMB substrate (Invitrogen) was added to each well and incubated for up to 2 minutes. 1 M HCl was added to stop the reaction. The absorbance was read at 450 nm using a Hidex sense microplate reader. Raw results were processed in excel. IC50 values were calculated in GraphPad using non-linear regression-[inhibitor] vs. normalized response.

2.2.7 Differential scanning fluorimetry

A solution of WWP2 HECT (2.5 μ M) and SYPRO orange (2.5 x from a 5000 x stock) in 50 mM Tris Cl, 150 mM NaCl and 3 mM DTT was prepared. 45 μ L of this mixture was pipetted into each well of a 96 well plate (MicroAmp Optical 96-well reaction plate or MicroAmp Fast Optical 96-well reaction plate). 5 μ L of compound was added to final concentration of 10 μ M, 0.1 % DMSO and the plate was sealed (MicroAmp Clear Adhesive Film). The plate was briefly centrifuged. The assay was then run on an ABI 7500 or ABI 7500 fast RT-PCR machine. The thermal profile was programmed to hold for 2 minutes at 25 $^{\circ}$ C, ramp to 70 $^{\circ}$ C at 0.5 $^{\circ}$ C per minute and then hold for a final 2 minutes at 25 $^{\circ}$ C. The melt temperature (T_m) was calculated by taking the first derivative of the resulting curve of fluorescence over temperature. Melt temperatures were calculated using the Thermo Fisher Protein Thermal Shift Software v1.3.

2.2.8 Hydrogen peroxide detection assay

Using 96 well plates (CorningTM CostarTM EIA), 45 μ L of ubiquitination buffer (described in section 2.1.6) and 5 μ L compound (1 % DMSO) were incubated for 30 minutes at room temperature. 50 μ L of HRP-PR (Horseradish peroxidase phenol red, 180 μ g mL⁻¹ HRP: 300 μ g mL⁻¹ phenol red) was added and incubated for 30 minutes at room temperature. 50 μ L of NaOH was added and the absorbance was read at 610 nm using a Hidex sense microplate reader.

2.2.9 Surface plasmon resonance

SPR experiments were carried out using a Biacore T200 instrument (GE Healthcare) on a Biacore Series S NTA sensor chip. A running buffer consisting of 10 mM HEPES pH 7.4, 150 mM NaCl, 0.005 % v/v tween 20 was used. The chip was generated with 0.5 mM nickel chloride and washed with 350 mM EDTA after the end of each cycle. The chip was regenerated for each cycle. SPR experiments were done in collaboration with Dr Clare Stevenson at the John Innes Centre, Norwich.

2.2.10 Statistics and repeats

All biochemical assay data is shown as a representative data set of at least two separate experiments. Each data point is shown as the mean of a triplicate. Error bars are calculated using standard deviation.

Chapter 3

Efforts towards the discovery and development of WWP2 inhibitors through biophysical and synthetic methods

3.1 Introduction

To design inhibitors for WWP2, there were two distinct routes that could be taken. As WWP2 belongs to the Nedd4 supergroup, which contain high homology, previous research into other inhibitors of this group could be utilised. The other approach is to initiate a high throughput screen with a selected technique which identifies protein ligand binding events. To use the first approach, there were two main compound "leads"; heclin (Mund et al., 2014) and "compound 3" (Kathman et al., 2015). Heclin, found through a phage display and peptide displacement assay, offers specificity towards HECT ligases over RING ligases (Mund et al., 2014). The presumed mechanism of action is that it causes a perturbation in the C-lobe of the HECT domain, resulting in exposure of the catalytic cysteine. This exposure is predicted to cause the cysteine residue to undergo chemical changes such as oxidation, leading to its loss of activity. There is no known binding site of heclin and no mechanistic explanation as to how it causes the conformational change. Although, the low solubility of heclin in aqueous solutions would present itself as a challenge for us to use in our subsequent biochemical and biophysical assays. Designing and synthesising analogues of heclin may improve the physiochemical properties, though there first needs to be a synthetic route which is amenable to analogue design.

The development of a Nedd4-1 covalent inhibitor was a significant discovery in the area of HECT ligase modulation. Not only did it report a covalent inhibitor with a crystal structure of the bound complex, the method of inhibition revealed an allosteric site involved in the building of poly ubiquitin chains (Kathman et al., 2015). Structure comparison of the Nedd4-1 HECT inhibitor complex and WWP2 HECT could be used to design potential compounds which would bind to WWP2 HECT. However, the lack of a non-catalytic site in WWP2 which the compound binds to Nedd4-1 is one of the reasons the compound is specific to Nedd4-1. It is unknown whether WWP2 HECT would still possess this allosteric site or

that it builds ubiquitin chains in this processive manner. Derivatives of indole 3-carbinol (Quirit et al., 2017) have been found to be inhibitors of Nedd4-1. These inhibitors were derived from "compound 3" and have been speculated to work in the same manner. Whether these compounds could be altered to target WWP2 remains to be seen. The second approach is to use high throughput screening to discover small molecules which either inhibit or bind to WWP2. There are a wide range of biochemical and biophysical techniques which can be used to identify inhibitors. Hits from the primary screen would be validated through an orthogonal assay, which has a different mechanism of action. The introduction to this chapter will briefly outline the biophysical techniques that were used to probe for WWP2 ligands.

3.1.1 Biophysical techniques

3.1.1.1 Differential scanning fluorimetry

Differential scanning fluorimetry (also known as thermal shift) is a common biophysical technique used for identifying protein ligands. The main principle is that the binding of a ligand increases the thermal stability of a protein. An increase in temperature causes protein tertiary structures to unfold, exposing hydrophobic surfaces. In the presence of a particular dye, the dye will bind to the hydrophobic surfaces. The binding of the dye causes a large increase in fluorescence. This fluorescence reading over temperature results in a sigmoidal curve. The mid point of this curve is coined the 'melt temperature' (T_m), and the binding can be evaluated by fitting fluorescence data into the Boltzmann equation or by taking the first derivative of the curve. If a ligand binds to the protein, there will be an increase in the melting temperature (Niesen et al., 2007). The assay can be run in 96 or 384 well format in RT-PCR machines, hence making it a desirable high throughput screening technique.

3.1.1.2 Surface plasmon resonance

Surface Plasmon resonance (SPR) is a highly sensitive technique used to detect ligand binding and calculate accurate K_d values. SPR works around the basis that a change in surface environment causes a change in the optical properties of polarised light. In SPR applications, the Kretschmann configuration is the common set up. This is where polarised light passes through a prism and hits the backside surface, the light reflects into the detector. At a particular incidence angle, known as the resonance angle, electrons on the backside surface absorb the light and resonate. These resonating electrons are known as surface plasmons. The result of these surface plasmons is a dip in intensity in the reflected beam and this dip can be monitored in the SPR reflection intensity. Surface plasmons are sensitive to their surrounding environment, and changes result in a change in angle of the intensity dip. Target ligands are immobilised onto the sensor chip, which acts as the backside surface in the Kretschmann set up. Analyte molecules are flowed over the sensor chip and binding events are detected as a function of SPR response where 1 resonance/ response unit (RU) equals 1° change in resonance angle (Patching, 2014). SPR can measure the SPR response over time meaning the kinetics of binding events can be calculated. The K_d is calculated by measuring the K_{on} and the K_{off} rates (Patching, 2014). SPR experiments are run on BiaCore machines and the sensor chip is specific to the desired method of capturing the ligand.

3.1.1.3 Saturation transfer difference NMR

NMR spectroscopy is an incredibly useful tool for studying molecules in solution. NMR can be used to gain structural and mechanistic insight into a wide range of molecular entities from small molecules to proteins. The versatility of NMR has been expanded to obtain information about the interactions between ligands and biological macromolecules. It has become an important step in the drug discovery process where NMR can be used as a screening purpose and to determine the bound orientation of the ligand. The saturation

transfer difference experiment is the primary NMR experiment exploited for these purposes. To understand the principles of STD NMR, the basics of NMR must first be explained. NMR uses the magnetic properties of particular nuclei, and nuclei in different structural environments have different magnetic properties. The quantum subatomic property spin determines the magnetic property of nuclei. Spin is a property of particles which have angular momentum, but does not describe rotational movement. There is no macromolecular equivalent to describe quantum spin and it is given a quantum mechanical description, the description of which is beyond the scope of this thesis. Nuclei with a spin quantum number greater than 0 are magnetic. The simplest nuclei is ^1H with a spin of $1/2$, hence the proton will be magnetic. When the proton is subjected to a strong magnetic external field, the nuclei will either align parallel or antiparallel to the field. This creates an energy transfer between the ground and excited state. When spin relaxation takes place, energy is transmitted at the same frequency as radio waves. These radio waves are processed to produce the typical 1D NMR spectrum. The energy transition is dependent on the proton magnetic field, which is affected by the chemical environment of the proton (Jacobsen, 2007). The STD experiment is based on the nuclear overhauser effect (NOE). The NOE is when the resonance intensity of a particular spin is affected by other spins in close space. This is known as the transfer of nuclear spin polarization through cross-relaxation. There are two spectrums obtained for the STD experiment. The on-resonance spectrum obtained by only irradiating the protein resonances, at 0-1 ppm. The off-resonance is the spectrum without irradiating the protein. During the on-resonance experiment, if the ligand is bound to the protein, it will receive saturation transfer via the NOE. This will cause the signal to disappear. Therefore, when the on-resonance experiment is taken from the off-resonance, the signals from bound ligands will remain (Viegas et al., 2011).

3.1.2 Experimental aims

WWP2 is an enticing novel target for therapeutics. The limited HECT ubiquitin ligase inhibitors discovered and none specifically for WWP2, makes searching for WWP2 inhibitors particularly interesting. This chapter sets out to discover, design, synthesise and evaluate novel WWP2 inhibitors. The layout of this chapter has three distinct sections. The first section focuses on employing a thermal shift experiment to screen for new WWP2 ligands. Following from the screen, approaches will be developed to validate binding of any hits, which is the focus of the second section. The third section of the chapter attempts to develop synthetic approaches for ligand hit and heclin. It is important to note that when starting this project, the only known general Nedd4 E3 ligase inhibitor was heclin. The methodology explored in this chapter combines to produce an experimental toolkit that can be used for the further development of WWP2 inhibitors.

3.2 Results

3.2.1 Utilising a thermal shift assay to identify WWP2 ligands

3.2.1.1 WWP2 HECT protein purification

To find novel ligands which could have the potential to be inhibitors, we chose the differential scanning fluorimetry technique. We focused on finding ligands for the HECT domain of WWP2. As the HECT domain contains the catalytic cysteine, we envisaged that any ligand binding event had the strongest chance of affecting its ubiquitination activity. The WWP2 construct reported by Wei Gong et al. (2015) where residues 486-865 correspond to the HECT domain, minus the last 5 residues, was chosen for thermal shift. As this was the truncation that produced a crystal structure, we predicted this would be the most stable truncation of HECT for thermal shift. The DNA sequence corresponding to this construct was cloned into a pET28a plasmid by In-Fusion cloning. The WWP2 HECT protein was expressed with an N-terminal 6xHis- tag. The protein was purified as described in Section 2.1.2.2, SDS page gel shown in Figure 3.1. Cleavage of the his tag is not necessary to run thermal shift experiments, so the protein was simply dialysed into the final stable buffer of 50 mM Tris-Cl pH 8.9, 150 mM NaCl and 3 mM DTT.

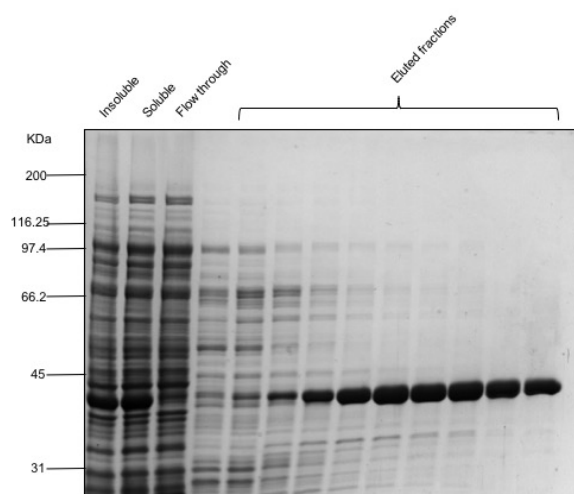


Fig. 3.1 -Nickel NTA purification of WWP2 HECT. The insoluble sample is post cell lysis. The soluble fraction is the clarified lysate after centrifugation. The flow through and wash samples is the sample collected after the clarified lysate is loaded on the column. The eluted fractions are 5 mL fractions collected after the high imidazole elution buffer is added to the column.

3.2.1.2 Thermal shift buffer screen

Once the protein was in the final Tris buffer, a small screen of conditions was performed. The protein was kept in the main Tris buffer for the screen due to stability issues in other buffers. Two protein concentrations (5 μ M and 2.5 μ M and three dye concentrations (5 x, 10 x and 20 x) were chosen to sample the quality of the melt profiles. Condition A and D both gave good thermal shift profiles, condition D was chosen as it gave a good profile and required less protein compared to condition A.

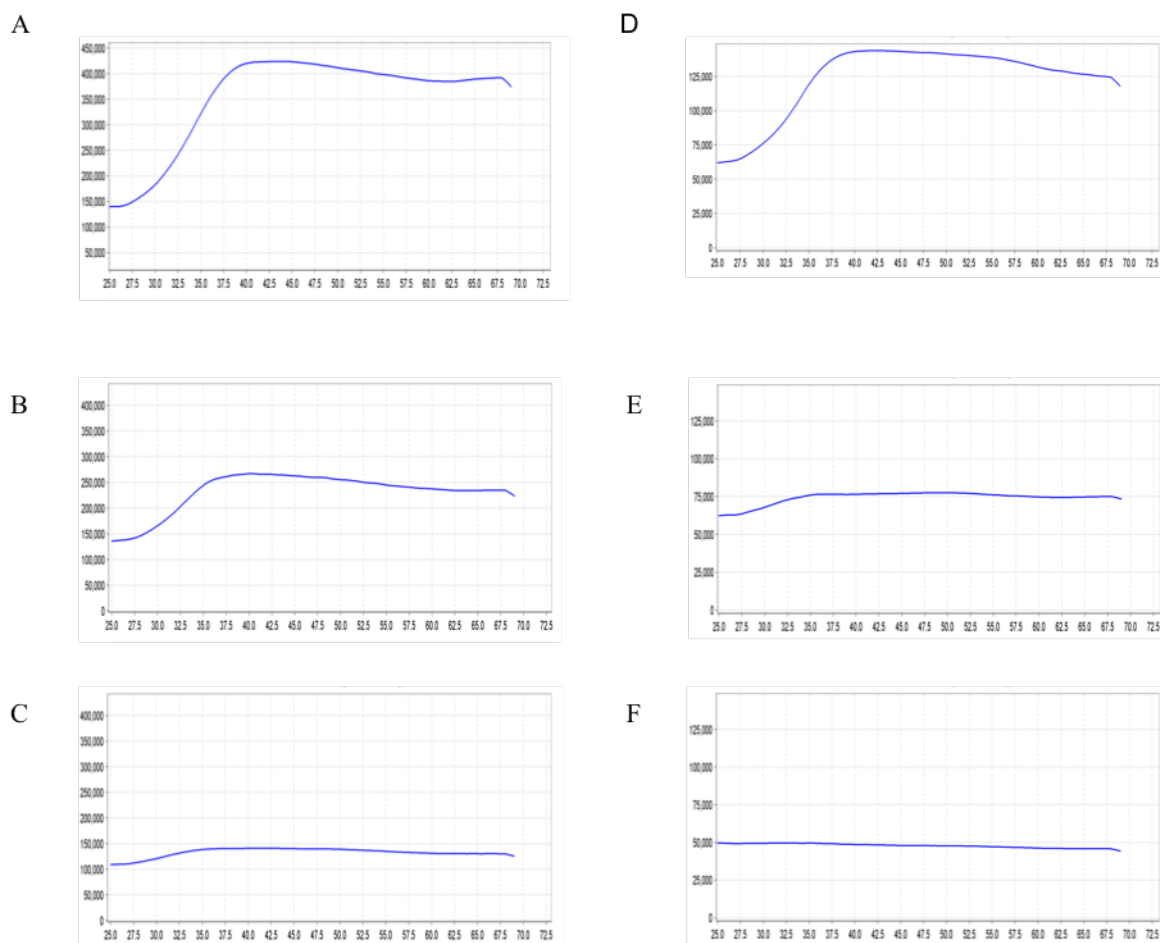


Fig. 3.2 -Condition screen for thermal shift (A) 5 μ M WWP2 HECT, 5 x SYRPO orange. (B) 5 μ M WWP2 HECT, 10 x SYRPO orange. (C) 5 μ M WWP2 HECT, 20 x SYPRO orange. (D) 2.5 μ M WWP2 HECT, 5 x SYPRO orange. (E) 2.5 μ M WWP2 HECT, 10 x SYPRO orange. (F) 2.5 μ M WWP2 HECT, 20 x SYPRO orange.

3.2.1.3 Compound screen

Next, we screened 1584 compounds from NCI diversity V and 419 compounds from NCI natural product set IV (Figure 3.3). The NCI diversity is a compound set derived from 140 000 compounds from the NIH National Cancer Institute Developmental Therapeutics Program repository. The large set of compounds is reduced to 1584 based on the criteria that the structure is rigid, 5 or few rotatable bonds, tendency to be planar, one or less chiral centres and with pharmacologically desirable features. The ‘diversity’ is derived from the

large number of possible pharmacophore permutations in the structures. The NCI natural product set consists of 419 natural product compounds from the NCI repository. The natural products were chosen based on structural diversity, purity, origin and availability of the compound. At this point we did not have a positive control for WWP2 HECT as there was no reported ligands. Therefore distinguishing hits from background noise would be difficult. Therefore in order to select compounds for triplicate experiments, we chose 2 °C. It is generally considered the greater the T_m, the higher the affinity of ligand (Mannhold et al., 2015). We wanted to choose a temperature shift that would be high enough to be distinguished from background noise and we felt a 2 °C shift would represent a significant distinguishable binding event. It would also limit the number of compound hits we could feasibly work with. The compounds in NCI diversity set V were distributed in 10 mM DMSO stocks. The NCI natural product set was also in 10 mM DMSO stocks but with 5 % added glycerol. Compound stocks were diluted to a final working concentration of 10 μM, 0.1 % DMSO for screening. Compounds were added to 2.5 μM WWP2 HECT on 96 well plates and underwent the thermal shift experiment described in section 2.1.6. DMSO, no DMSO and no protein controls were run on each plate and the curves were inspected to confirm a successful run. Compounds with a melt temperature which represented a 2 °C shift above the average, were visually inspected and those which gave a typical curve were identified as preliminary hits. Of the diversity set, 22 compounds were picked out to have a temperature shift equal or greater than 2 °C. These compounds were run in triplicate and 5 compounds showed a consistent temperature shift (Figure 3.3 and Table 3.1). NSC601359 showed the greatest shift melt temperature with a 2.65 °C increase. The five compounds are fairly structurally diverse apart from NSC601359, NSC33005, NSC35676 and NSC215718 being phenolic compounds. NSC601359 and NSC215718 are the most similar compounds with hydroxyl substituted naphthalene groups and a bromine containing aromatic ring. NSC35676 is a natural product known as Purpurogallin. From the 419 natural products, 11 compounds

produced a significant temperature shift. Purpurogallin was also present in the natural product compound library and was again identified as a hit. This compound was not included in the 11 identified hits as the follow up experiment had already been carried out. Triplicate runs reduced this number down to 6 natural products (Figure 3.3 and Table 3.2). Five of the six triplicate validated natural products are phenolic.

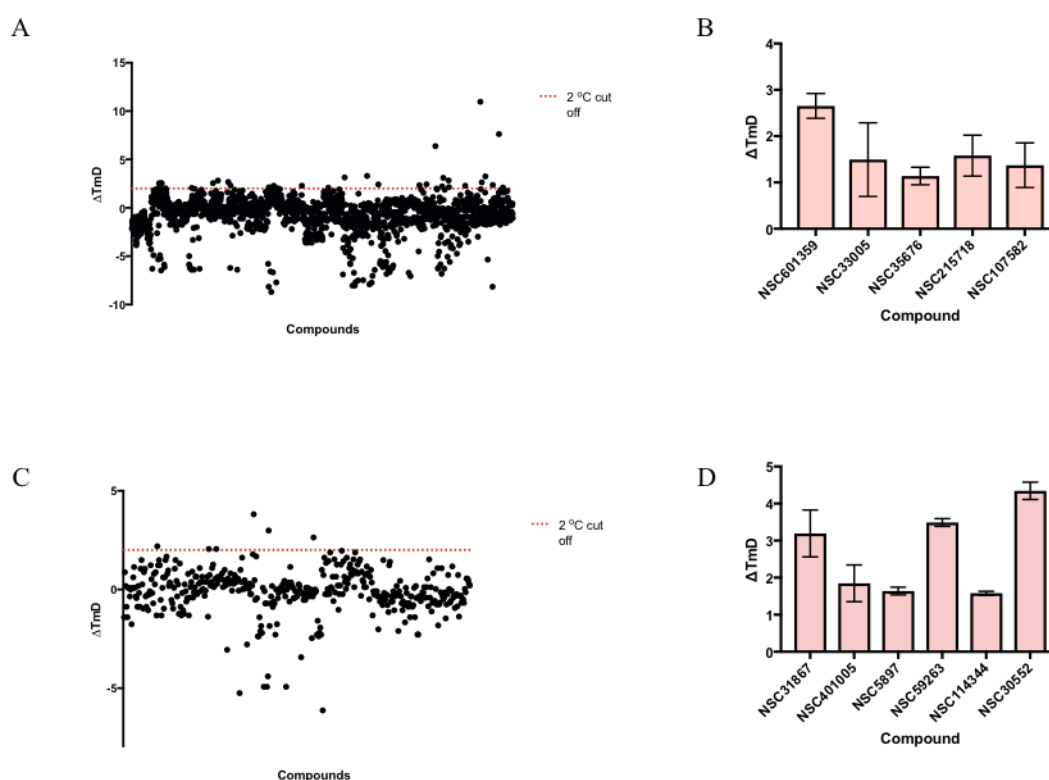


Fig. 3.3 -Thermal shift screen (A) Diversity set V thermal shift screen. 1584 compounds were screened single shot at 10 μ M in 0.1 % DMSO. Controls of + DMSO, - DMSO and NPC (no protein control) were run on each plate. The control T_m was the average T_m of the + DMSO control. ΔT_m of the screened compounds were calculated by the difference between single compound T_m and the control T_m . (B) Diversity set hit compound triplicates (C) Natural product set IV screen. The screen was carried out in the same way as described for the diversity set screen, but with 0.05 % glycerol added in the control well solution. (D) Natural product set hit compound triplicates

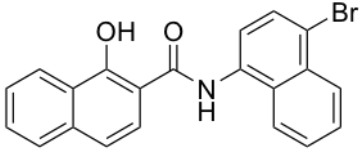
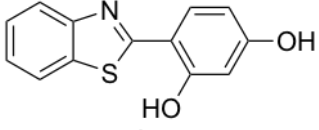
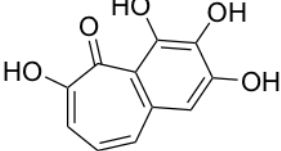
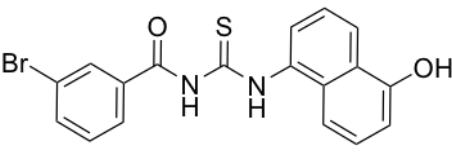
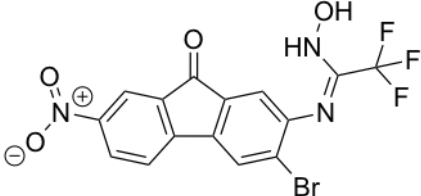
NSC number	Compound structure	$\Delta T_m D$ ($^{\circ}C$)
601359		2.65
33005		1.49
35676		1.14
215718		1.58
107582		1.37

Table 3.1 -Structures and $\Delta T_m D$ of Diversity set hits

3.2.1.4 Biological testing of hit compounds

The compounds were tested for whether they had any biological activity against WWP2. The WWP2 autoubiquitination assay was employed. This mimics the ubiquitination cascade *in vitro*. GST WWP2 is anchored down by glutathione coated plates, and then incubated in the presence of E1, E2, flag tagged ubiquitin and ATP. After subsequent washings, flag tagged ubiquitin can be recognised by the anti-Flag antibody. This antibody also had a horseradish peroxidase (HRP) tag which acts a substrate for 3, 3', 5, 5'- tetramethylbenzidine (TMB). This reaction causes a colour change of the TMB of clear to blue. This colour change can be measured quantitatively by stopping the reaction with HCl and reading the absorbance at 450 nm. Compounds were added to the reaction at the concentrations 1 μM , 5 μM , 10 μM ,

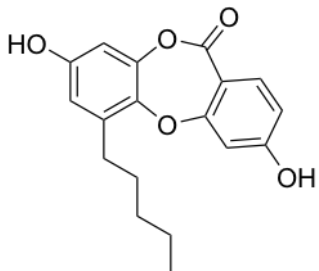
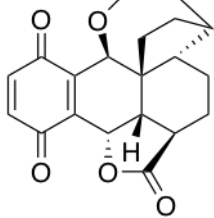
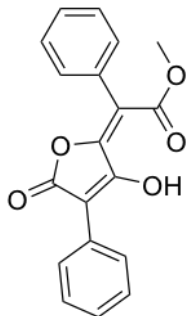
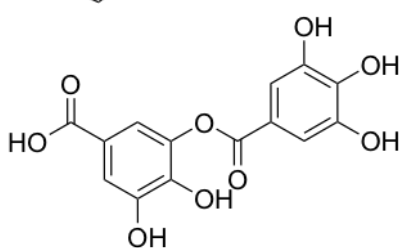
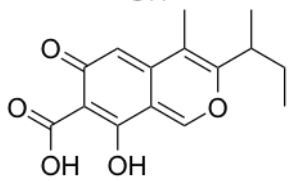
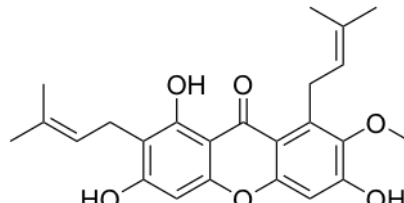
NSC number	Compound structure	ΔTmD °C
31867		3.19
401005		1.85
5897		1.64
59263		3.49
114344		1.58
30552		4.34

Table 3.2 -Structures and ΔTmD of Natural product set hits

25 μM , 50 μM and 100 μM . NSC601359, NSC33005, NSC215718, NSC107582 from the diversity set showed no inhibition and an IC₅₀ could not be calculated (Table 3.3 and Figure 3.4). Purpurogallin, the compound which was picked out as a hit in both the diversity and natural product sets inhibited WWP2 with an approximate IC₅₀ of 97.57 μM . During the time of these experiments taking place the same compound libraries were being screened by the autoubiquitination biochemical assay in the Chantry lab (carried out by Jess Watt). Pleurotin came out as a hit in the biochemical screen and was reported with an IC₅₀ of 2.67 μM (Jess Watt 2018) (data not shown). Digallic acid inhibited WWP2 with an IC₅₀ of 187.9 μM .

Compound	IC ₅₀ μM
NSC601359	N/A
NSC33005	N/A
NSC215718	N/A
NSC107582	N/A
NSC35676	97.57
NSC31867	N/A
NSC401005	2.67
NSC5897	N/A
NSC59263	187.9

Table 3.3 -IC₅₀ values of thermal shift hit compounds for the WWP2 autoubiquitination assay

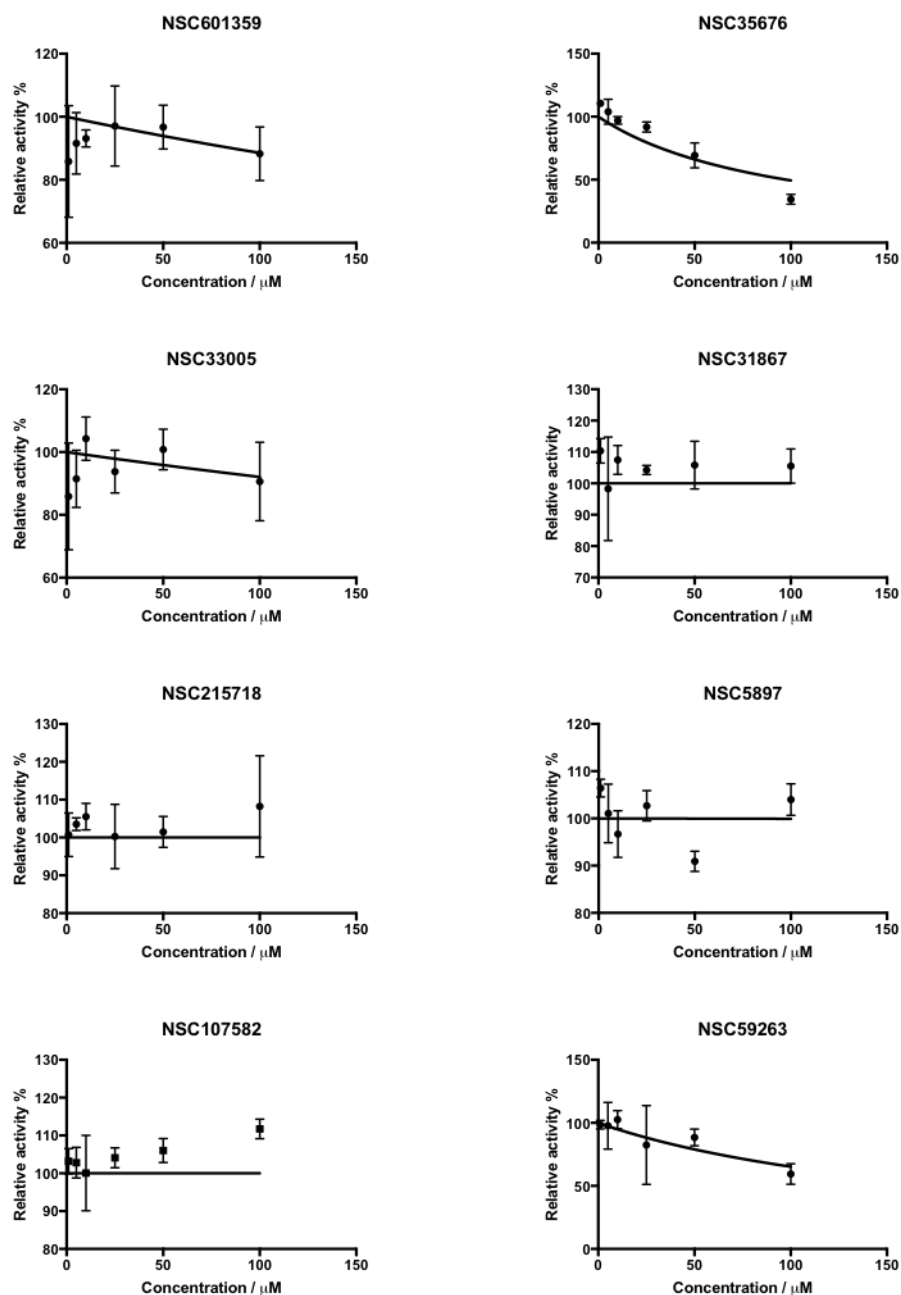


Fig. 3.4 -Thermal shift hit compounds IC₅₀ curves for the WWP2 autoubiquitination assay. Thermal shift IC₅₀ curves corresponding to the values reported in Table 3.3. Only NSC35676 and NSC59263 show a concentration dependent response

3.2.2 Orthogonal approaches to validate ligands

At this point we chose to narrow down the hit compounds. NSC35676, NSC401005 and NSC59263 all showed inhibitory activity and therefore were selected for further analysis. The four other hits from the diversity set were also selected. This was because although they showed no inhibitory effect, they could still be WWP2 ligands. The structural composition of these hits were simpler than the hit from the natural products, so we envisaged any validated ligands could be manipulated chemically to induce inhibition further down the pipeline. It is commonplace in drug discovery to validate hits through an orthogonal biophysical assay. The next part of this chapter explores two biophysical techniques; STD NMR and SPR, to validate selected ligands.

3.2.2.1 Orthogonal validation by STD NMR

The first orthogonal technique we tried was STD NMR. The WWP2 HECT protein was prepared in 25 mM Tris-HCl pH 8.9, 100 mM NaCl and 1.0 mM DTT. STD NMR experiments were run in collaboration with Dr Jesus Angulo, Dr Sam Walpole and Dr Serena Monaco. For the NMR sample, there was 20 μ M protein and 100 μ M compound in 1 % deuterated DMSO, 10 % D₂O. The results for NSC601359, NSC33005, NSC215718, NSC107582 were deemed inconclusive. The STD results suggested they might be binding but the signal was too weak to say with confidence that the compounds were ligands. NSC35676 and NSC401005 showed a positive result and were confirmed as binders. The STD spectrum of NSC35676 is shown in Figure 3.5.

The spectrum of NSC401005, as shown in Figure 3.6, in the presence of the protein had some of the peaks moving, and /or disappearing. Due to this, we led to believe the compound was degrading or undergoing a reaction. Whilst it would have been interesting to determine whether the compound was undergoing chemical changes due to the conditions of the buffer

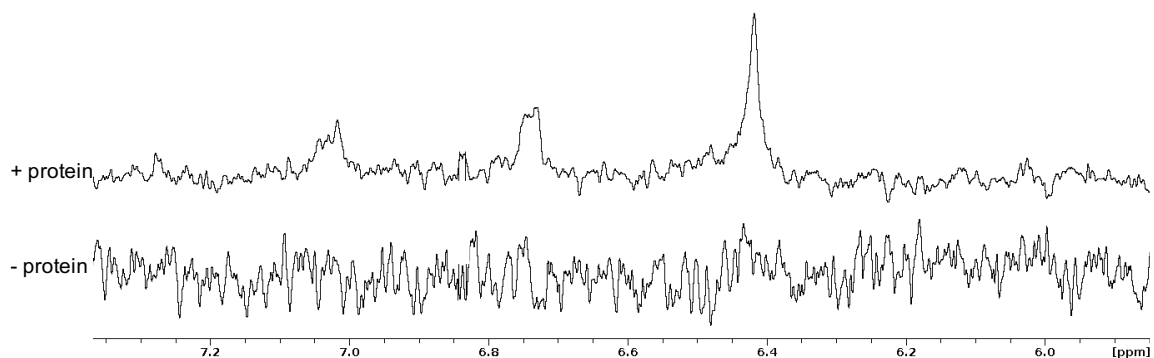


Fig. 3.5 -STD spectrum of NSC35676.

or whether the changes were protein induced, limited access to the solid and the 800 MHz NMR machine meant further experiments were not pursued.

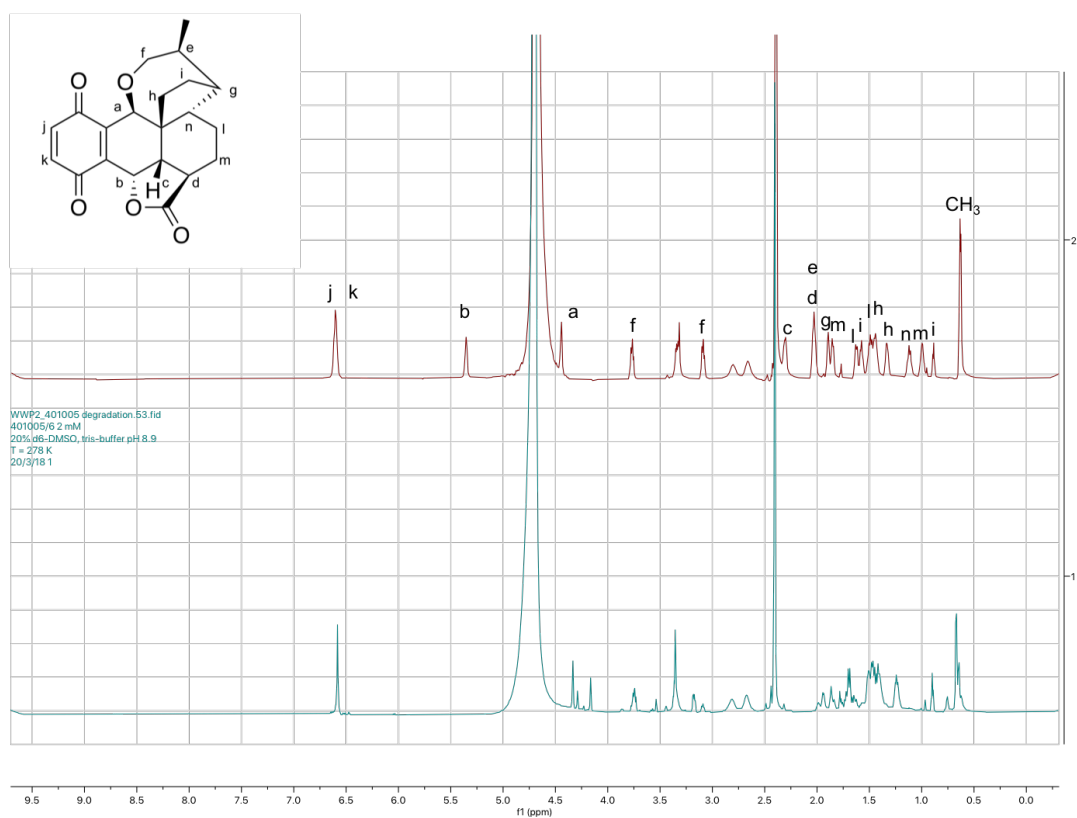


Fig. 3.6 -NMR spectrum showing the degradation of NSC401005 in the presence of WWP2 HECT. Assignment was done by Dr Serena Monaco.

3.2.2.2 Using surface plasmon resonance to investigate binding to WWP2

To investigate possible binding events to WWP2 we chose to see whether surface plasmon resonance (SPR) could be used as an alternative biophysical technique. The output of the SPR is via a direct response to binding, as opposed to thermal shift which is indirect. SPR was an appealing technique to detect binding to WWP2 because it could be used as a confirmatory technique, work out K_d and $K_{on/off}$ rate and the nature of the experimental set up means it could be adapted to high throughput screening.

3.2.2.3 Purification of WWP2 FL

The full length isoform of WWP2 (WWP2-FL) was chosen for SPR. Compounds which were discovered in the biochemical assay could be also then be validated using the autoubiquitination assay as the full length construct is used. SPR chips can anchor biological macromolecules by many bioconjugation methods, but one of the most common is using a NTA chip. The NTA chip is coated with nickel, which can then bind to His tagged proteins. WWP2- FL was expressed in the CodonPlus RP cell line after 1 mM IPTG was added and incubated at 16 °C overnight. The protein was extracted using the Ni NTA procedure described in section 2.1.2.3. Gel filtration was used to gain a purer sample following elution from the Ni NTA column. Two prominent bands co eluted around 98 kDa and 70 kDa. WWP2- FL Curiously, the 70 kDa band eluted first, where the largest protein should elute first from the gel filtration column. Samples containing both bands were pooled and run through the Ni NTA column to see whether both samples were His tagged. Both samples bound to the column and eluted in imidazole buffer suggesting both proteins were His tagged (Figure 3.7). We were fairly confident that the 98 kDa fraction was WWP2 FL, though the 70 kDa fraction was still unknown. As at least 50 % of the sample was our target protein, and the other 50 % a probably variation on our protein, we decided to trial SPR with this

sample. The protein was stored at an assumed concentration of 3.97 μM in 25 mM Tris-Cl pH 8.0, 250 mM NaCl, and 5 mM DTT.

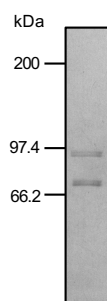


Fig. 3.7 WWP2 FL protein purification

3.2.2.4 WWP2 SPR experiments

The following work was done in collaboration with Dr Clare Stevenson at JIC, Norwich. We first tried to bind the WWP2-FL sample to the Ni NTA Biacore chip (Figure 3.8). All SPR experiments used the running buffer consisting of 0.01 M HEPES pH 7.4, 0.15 M NaCl and 0.005 % v/v tween20. WWP2 FL at a concentration of 40 $\mu\text{g mL}^{-1}$ was run over the chip, however no response was seen, indicating there was no binding. As we also had the His-tagged WWP2 HECT construct available, we tried binding this protein at 64 $\mu\text{g mL}^{-1}$ as an alternative. A clear response was measured, indicating the successful binding of WWP2 HECT. We therefore went forward for binding experiments with the HECT domain instead of WWP2 FL.

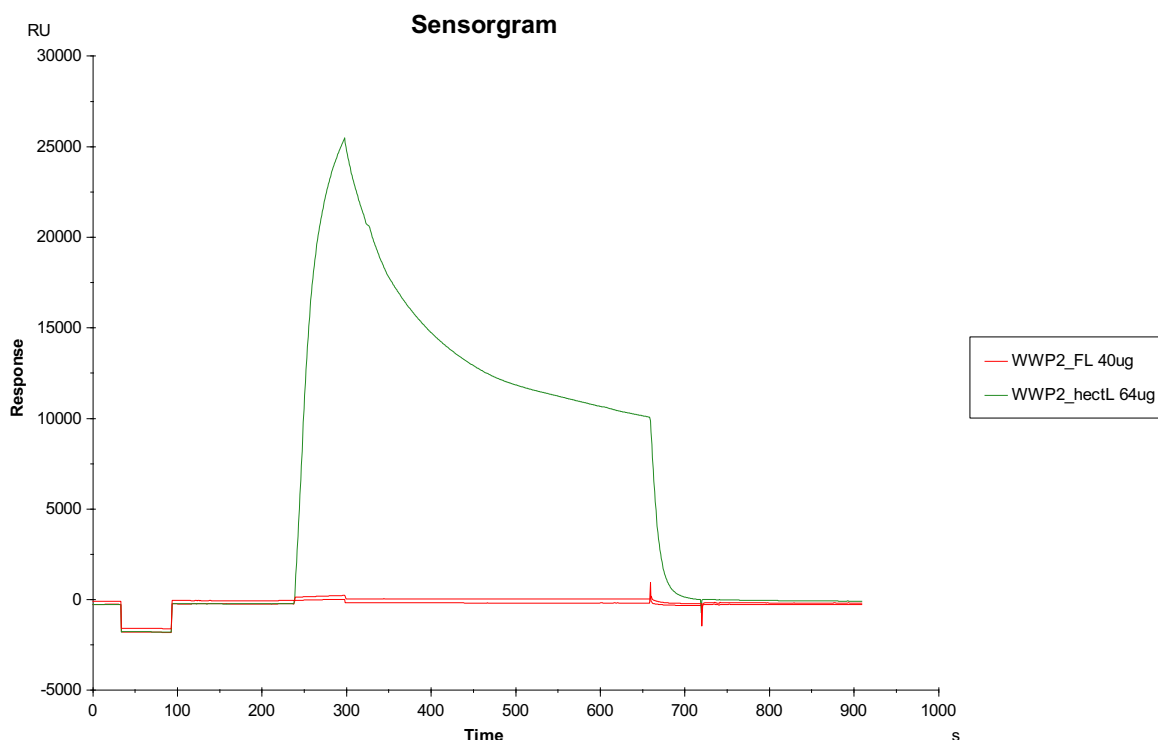


Fig. 3.8 SPR sensorgram of WWP2 fl and WWP2 HECT

3.2.2.5 Detection of ligand binding to WWP2 by SPR

Three already validated WWP2 HECT ligands were chosen for SPR analysis; NSC35676, NSC2883887 and NSC2805. NSC2883887 and NSC2805 were WWP2 inhibitors identified in our lab and reported by Watt et al. (2018). Each compound was flowed over the chip in a concentration of 1 μ M in 0.002 % DMSO. The chip was regenerated with nickel and protein between each run. A control of 0.002 % was used as a negative control. All three compounds presented evidence of binding in the sensorgram (Figure 3.9).

The next step was to measure the response of the ligands against the response of non-binding compounds. This was to ensure the signal from a binding event would be distinguishable from background noise during screening. We chose the same compounds as previously mentioned, at 1 μ M and 10 μ M concentrations. Eight random compounds were chosen from the NCI diversity set V, which were not WWP2 ligands. Each random compound was tested at 10 μ M. As the SPR response is dependent on the change in mass on the SPR chip surface,

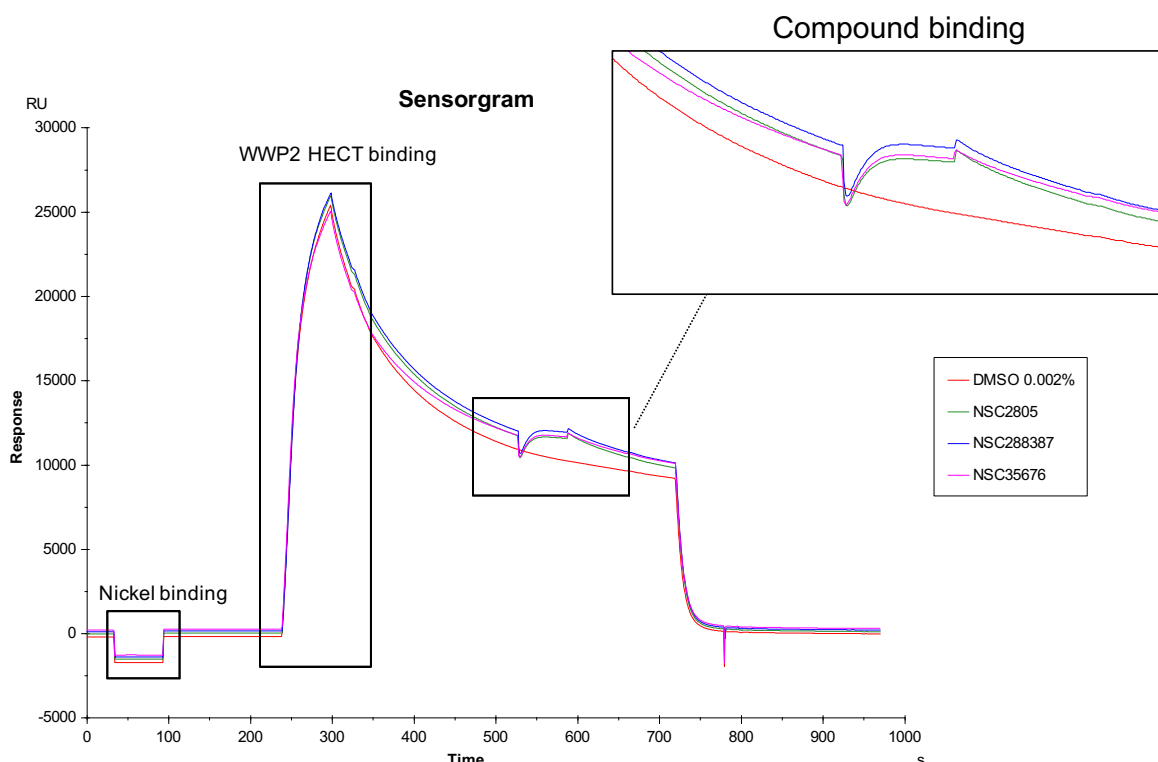


Fig. 3.9 -Compound binding to WWP2 HECT

the MW of the compounds is taken into account. The raw response units were converted to the R-max by the following equation. Where $MW_{analyte}$ is the molecular weight of the small molecule, MW_{ligand} is the molecular weight of WWP2 HECT, R_{ligand} is the response from the protein and V_{rec} is the stoichiometry of the binding event. In this case, the exact stoichiometry is unknown so is presumed as 1. Each value was normalised by removing the average buffer value.

$$R_{max} = \left(\frac{MW_{analyte}}{MW_{ligand}} \right) \times R_{ligand} \times V_{rec}$$

Fig. 3.10 -R-max equation

The response from the random eight known non binders were taken as the background noise level. Significant values of the known ligands were calculated using a two tailed t-test against the average of the non-binders. NSC2805 showed very good significance with both

data sets ≤ 0.01 . NSC35657 at 1 μM showed good significance at the 0.05 level but at 10 μM showed none. NSC288387 did not show any significant response. It was clear the noise level was too high to reliably detect all possible binding events in a compound screen (Figure 3.11). Follow up experiments were attempted to improve the noise level, including adding DMSO into the running buffer and doing dose dependent experiments with more data points for each ligand. However, no interpretable data was produced. We felt SPR could be used as a technique to detect WWP2 ligand binding, but there was significant amount of optimisation required. Due to lack of resource and the limited time scope for this research, this technique was not pursued further.

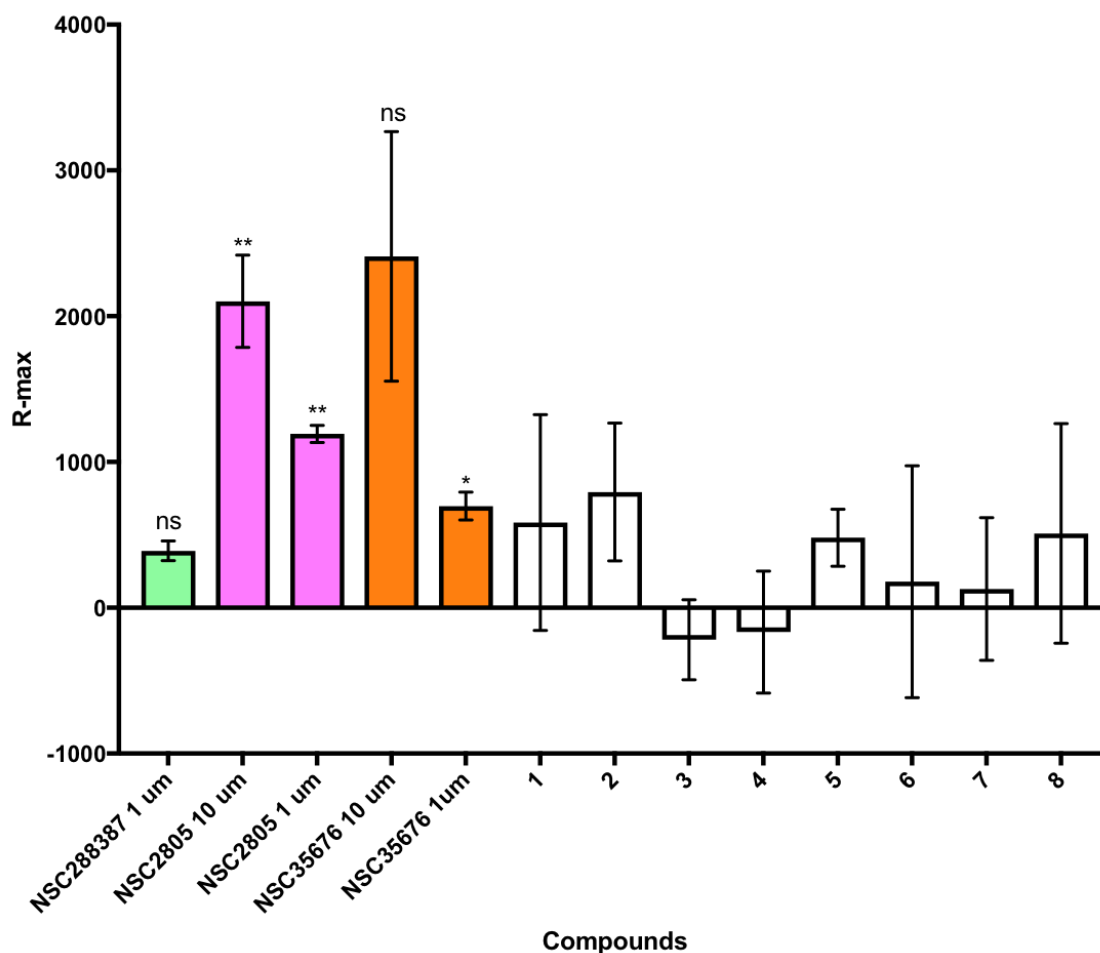


Fig. 3.11 R-max values of compound binding to WWP2 HECT

3.2.3 Synthetic approaches

The next part of this chapter introduces a synthetic chemistry approach to develop WWP2 inhibitors. Initially a route to synthesise heclin is explored, and then there is some further development on NSC35676, an inhibitor discovered in Section 3.2.1.3.

3.2.3.1 Heclin synthesis

Heclin is the compound identified as HECT specific ubiquitin ligase inhibitor. At the time of starting this research, it was the only compound reported in the literature as a general inhibitor for HECT ligases. Interestingly, when researching the compound in the literature, there was no reported synthesis. Given the relatively simple chemical structure of heclin, we felt there was scope for a short synthesis to be devised which could later be used to gain access to a wide variety of analogues (Figure 3.12). Looking at the structure of heclin, the most obvious bond to break was the amide bond. The corresponding amine would be the commercially available 4-acetylaniline. The carboxylic acid can be formed by a Knoevenagel condensation with Doebner modification reaction from 5-methylfurfural. The 5-methylfurfural can be synthesised from the commercially available 2-ethyl furan via a Vilsmeier- Haak reaction.

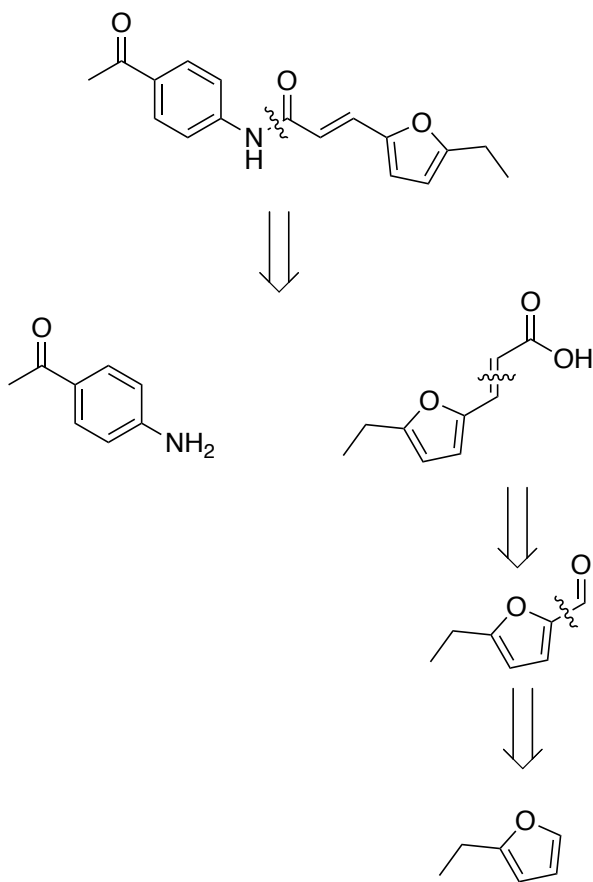


Fig. 3.12 -Scheme 1. Heclin retrosynthesis.

3.2.3.2 Vilsmeier-Haak reaction

The Vilsmeier-Haak reaction involved the formylation of electron rich aromatic compounds using a Vilsmeier reagent. The Vilsmeier reagent is a iminium salt in equilibrium formed from *N,N*-dimethylformamide (DMF) and an acid chloride. The reaction favoured the desired position of formylation on 2-ethylfuran as the Vilsmeier-Haak reaction is regioselective, favouring the less hindered *para* position of substituted rings. The Vilsmeier-Haak reagent was formed *in situ* by reacting DMF with phosphorous oxychloride. 2-Ethyl furan (**1**) was added and the reaction was left for 2 hours. Following the work up, the crude was purified by fractional distillation to give **2** in a 79 % yield.

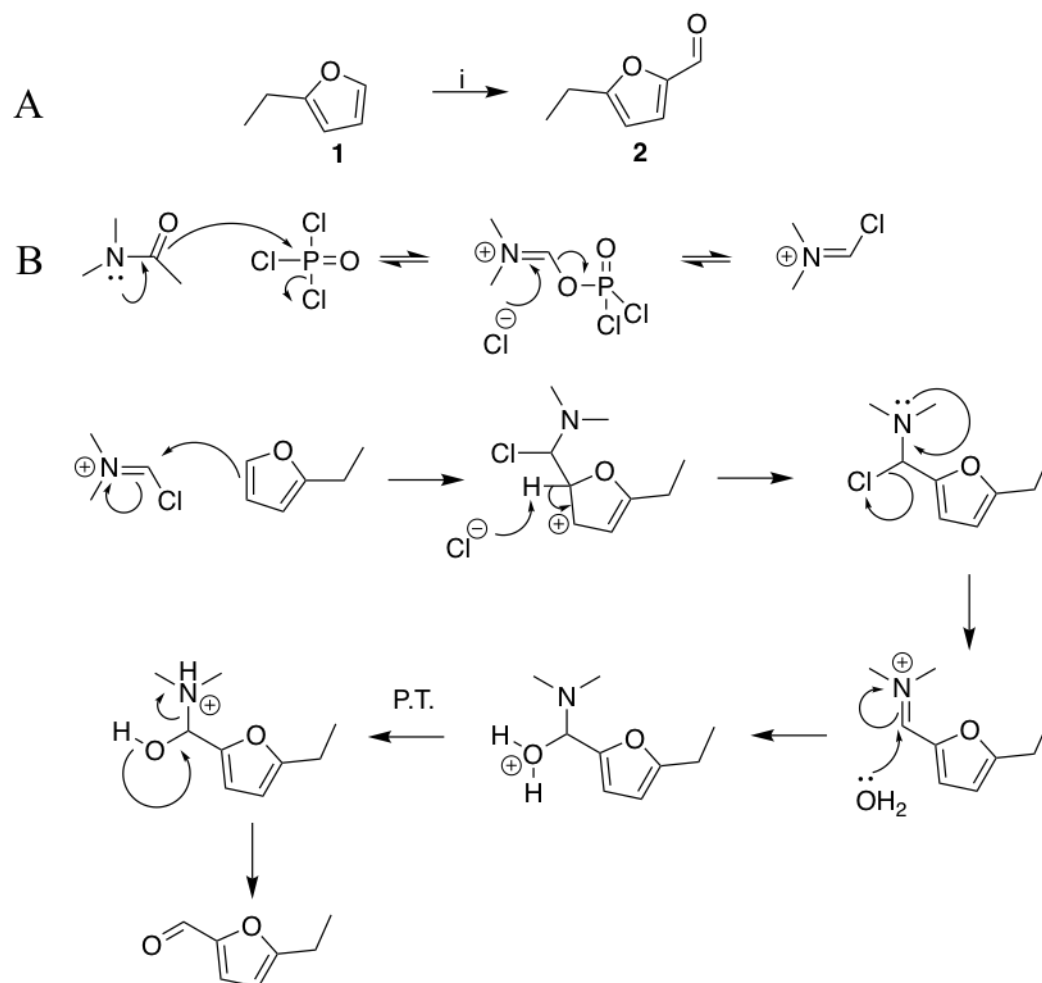


Fig. 3.13 -Scheme 2. The Vilsmaier Haak reaction. (A) Reaction conditions: (i) POCl₃, Na₂CO₃, DMF, 0 °C, 79 %. (B) Vilsmeier-Haak reaction mechanism.

3.2.3.3 Knoevenegal condensation Doebner modification

The Knoevenegal condensation is the nucleophilic addition to an aldehyde to form α , β -unsaturated compounds. The Doebner modification is the use of malonic acid instead of malonate, and under refluxing pyridine leads to decarboxylation. Compound **2** was reacted with piperidine and malonic acid in pyridine and upon work up the α , β -unsaturated carboxylic acid **4** was isolated in 47 % yield without the need for further purification.

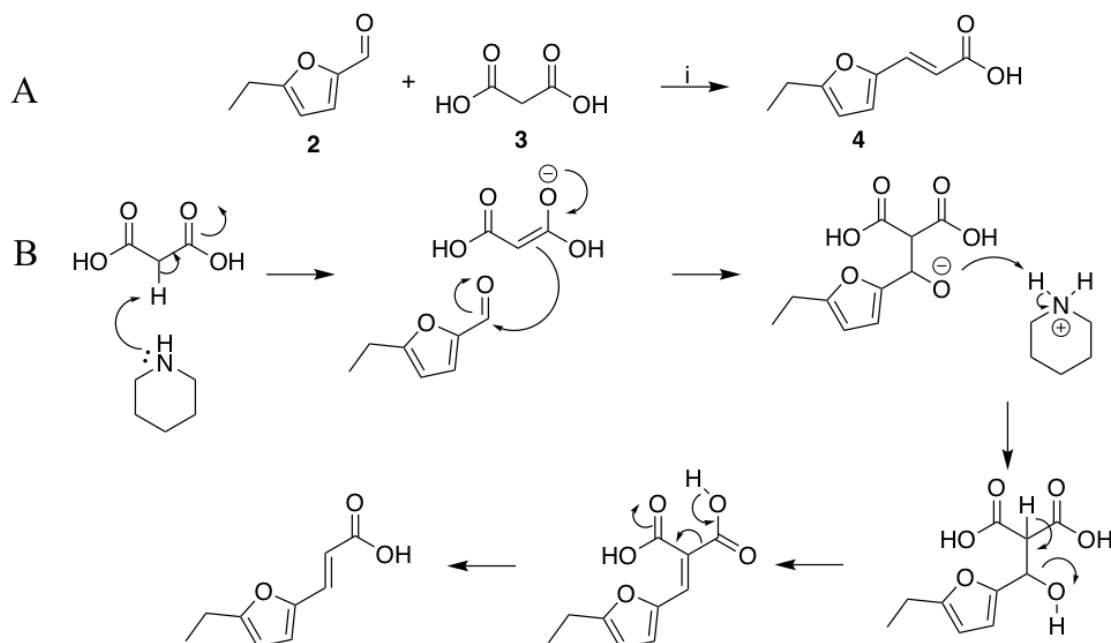


Fig. 3.14 -Scheme 3 (A) Reaction conditions: (i) Piperidine, malonic acid, pyridine, 47 %. (B) Knoevenegal condensation mechanism.

3.2.3.4 Amide bond formation

Amide bond formation between (**4**) and 4-acetylaniline was attempted using DCC and DMAP in dichloromethane at room temperature. Crude NMR suggested formation of the product. However upon purification by column chromatography, the NMR and MS attributed the structure to be of the *O*-acylisourea intermediate (**5**). This suggested the 4-acetyl aniline was not reacting with the *O*-acylisourea intermediate. To our delight, simply heating the reaction to 50 °C resulted in the formation of **7**. The product was purified by column chromatography in 26 % yield (scheme 4).

3.2.3.5 Synthesis of the purpurogallin scaffold

From the thermal shift hits, there were only two which were orthogonally validated ligands with inhibitory activity; NSC35676 and NSC401005. Chemically, NSC401005 is complex with six stereocentres and a six ringed compact cyclic framework. The racemic total synthesis has been completed in 26 steps (Hart and Huang, 1988); there is still no enantioselective

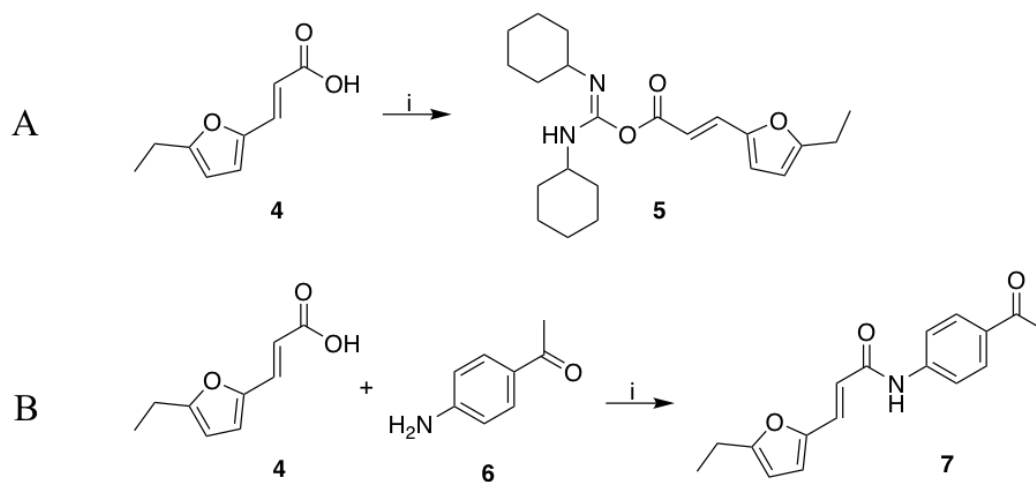


Fig. 3.15 -Scheme 4 (A) Conversion to *O*-acylisourea. Reaction condition:(i) DMAP, DCC, 4- acetyl aniline, DCM, room temperature, 39 %. (B) Conversion to heclin. Reaction conditions: (i) DMAP, DCC, 4-acetyl aniline, DCM, 50 °C, 26 %.

synthesis. Due to the formidable synthetic challenges faced in any optimisation and the ambiguous potential structural change upon binding, we chose not to advance this compound synthetically. Purpurogallin on the other hand is chemically simpler. It is a benzenetriol fused to a hydroxy ketone substituted seven membered ring called a tropolone. It is possible to synthesise purpurogallin scaffold in a single step using a horseradish peroxidase catalysed oxidation with pyrogallol and catechol. We envisaged this could be exploited to quickly make analogues to see whether any SAR emerges.

We wanted to attempt to leave one of the hydroxyls free for further functionalisation, hence we needed to add protecting groups. The hydroxyl groups on phenolic compounds can undergo methyl etherification. The methyl ether group can act as a protecting group for the hydroxyl. It would also change the group from a possible hydrogen bond donor/acceptor to just an acceptor. We chose this group as the resulting compound will not be dramatically altered and it would still be interesting to biologically test. Starting from 3-methoxycatechol and methyl 3,4,5-trihydroxybenzoate, the reaction conditions of 3 % H_2O_2 , HRP in citrate buffer with 5 % acetone gave the resulting purpurogallin analogue crude. Initially purification by column chromatography was attempted, however the product streaked and stuck to the

stationary phase, resulting in poor separation and recovery. Recrystallisation in ethyl acetate, acetone, methanol and ethanol were carried out in order to find an alternative to column chromatography. Success was achieved by recrystallisation in ethanol (scheme 5).

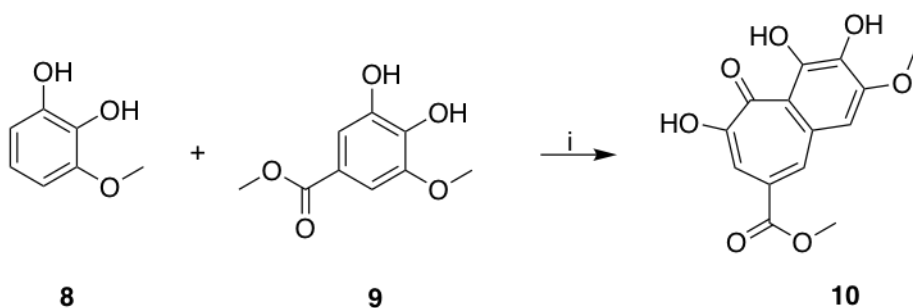


Fig. 3.16 -Scheme 5. Reaction conditions: 3 % H_2O_2 , HRP, phosphate citrate buffer pH5, 20 % acetone, 1 hour, 50 % yield

Based on the condition reported in Cheng et al., two of the hydroxyl groups in compound **10** were methylated using methyl iodide, leaving one hydroxyl group free (scheme 6). A NOESY experiment was run to determine which hydroxyls had been methylated.

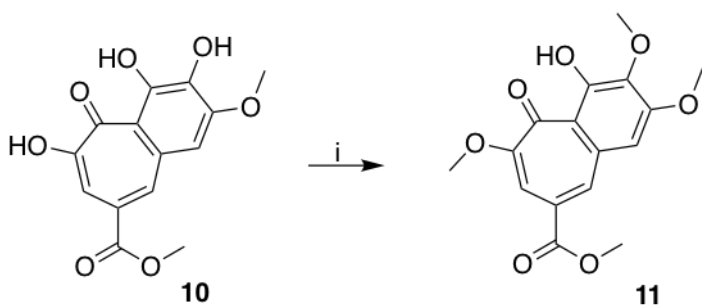


Fig. 3.17 -Scheme 6. Reaction conditions: Methyl iodide, potassium carbonate in DMF, overnight, 64 %.

3.2.3.6 Purpurogallin analogue testing

The purpurogallin analogues **10** and **11** were tested by the same method as described previously in Section 3.2.1.4. Concentration dependent assays were set up at 5 μM , 10 μM , 25 μM , 50 μM and 500 μM . The higher concentration was chosen as some of the compounds may have reduced inhibition than IC₅₀ of purpurogallin which was 97.57 μM . Purpurogallin was repeated and compared with the analogues at the new range of concentrations. The IC₅₀ curves are shown in Figure 3.18. Compound **11** showed no inhibition, and compound **10** showed an increase in potency by approximately two fold to 51.65 μM . NSC35676 produced an IC₅₀ of 162 μM , whilst this is lower than previously measured, there is still a comparative increase in potency for **10**. There was visible precipitation for **10** when preparing the concentration solutions.

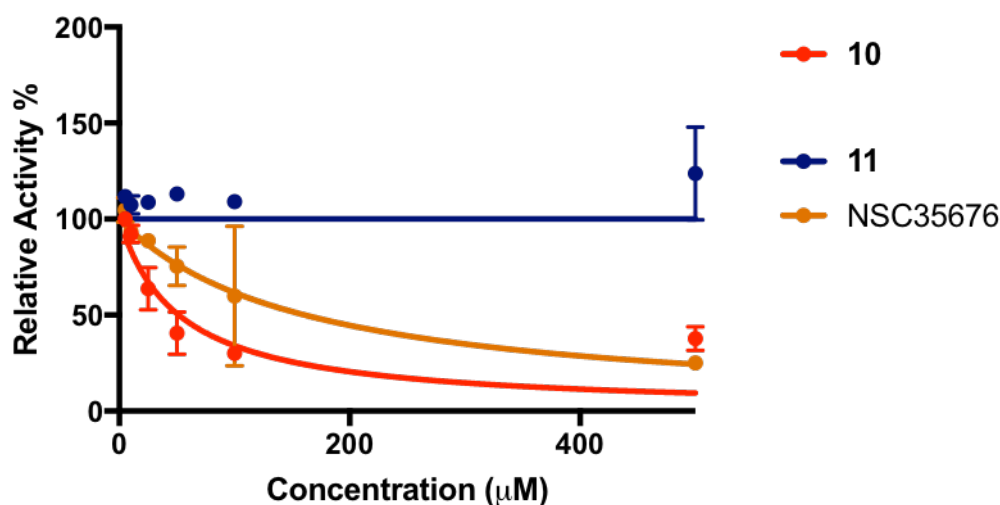


Fig. 3.18 -Purpurogallin analogue testing against WWP2 autoubiquitination

The purpurogallin analogues were tested for Nedd4 autoubiquitination inhibition. The compounds were tested at 1 μM , 5 μM , 10 μM , 25 μM , 50 μM and 100 μM in 1 % DMSO. NSC35676 gave an IC₅₀ of 37.78 μM which is lower than when tested against WWP2. As with WWP2, compound **10** produced a comparatively lower IC₅₀ than NSC35676 at 4.21

μM . The same pattern seen with WWP2 is also observed for compound **11**, which showed no inhibition.

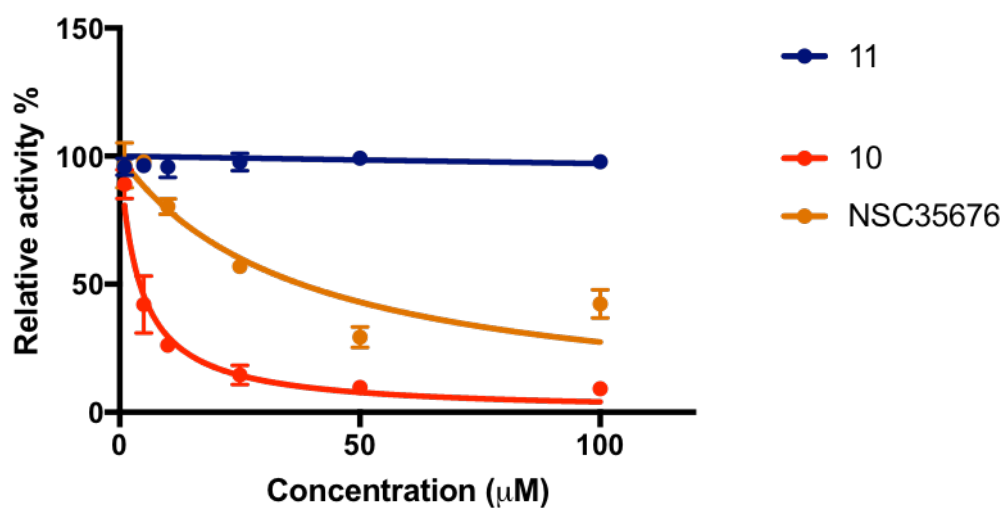


Fig. 3.19 -Purpurogallin analogue testing against Nedd4 autoubiquitination

3.3 Discussion

Rational drug design is the principle of selecting a biological target which has been validated as a crucial component in the molecular pathways of a chosen disease. The target protein must be cloned, expressed and purified in sufficient amounts for the multitude of biophysical and biochemical experiments needed to identify specific inhibitors. Choosing the protein construct to express is key, as proteins can contain multi domains, and targeting some of those domains may not be necessary in inducing inhibition. Particular constructs may be more stable than others, different expression levels and some may have differential stability depending on the technique used. Once the protein has been successfully purified, a biophysical technique is used to perform compound screening. Hits from the screening are further investigated and characterised. These compounds are then optimised to have the desired drug like properties. WWP2 has been proven to be an interesting therapeutic target for cancer. Taking WWP2 through the drug design process could lead to potential cancer therapeutics as well as molecular probes for WWP2. Little work has been previously published on screening for small molecule inhibitors for HECT proteins, and none on WWP2 (at the time of starting this research). The aim of this chapter was to set up a screening platform for WWP2 and discover ligands which could be used as a chemical scaffold for further optimisation. The chapter also sets out to report the first synthesis of the first HECT inhibitor in a way that could give rapid access to a wide variety of analogues. Three key biophysical techniques were trialled; thermal shift, STD NMR and SPR. From a practical sense, thermal shift was the easiest method to be explored for high throughput screening. Previous work in the Chantry lab by Dr Lloyd Whal (Whal 2015) set a precedent in the challenges in expressing and purifying sufficient amounts of biologically active isoforms of WWP2. We therefore chose to just focus on the HECT domain, for which the crystal structure was published at the time of commencing this work. The HECT domain is still

biologically relevant as it is the domain within WWP2 which holds the catalytic capability. Hence targeting this domain should inhibit WWP2 activity. The HECT domain proved to be easily expressed and purified as a His tagged protein. Brief optimisation resulted in thermal shift curves which were deemed suitable for compound screening. Not having a positive control at this point was a significant issue, as we could not determine the level of signal to noise for the assay. Heclin had been reported as an HECT inhibitor, the proposed mechanism of action was through forcing a conformation change in the structure of HECT which results in oxidising the catalytic cysteine, removing the cysteine of its catalytic ability. The authors did not confirm that heclin was a ligand for WWP2. When tried in the thermal shift assay (results not shown) there was no shift detected. Therefore, we could not use the only potential positive control. There remains speculation as to whether heclin is a true inhibitor for WWP2, or that all HECT ligases can be inhibited by this mechanism of action. Screening of the NCI diversity set V and natural product set IV resulted in the identification of 11 hit compounds. Apart from all containing aromatic rings, they were fairly structurally diverse. The biological testing of these compounds resulted in only NSC35676, NSC401005 and NSC59263 to be inhibiting WWP2. We still envisaged the other compounds could be ligands. Finding new ubiquitin ligase ligands is currently of great interest to due the emergence of PROTACs. With the need to orthogonally validate the inhibitors and ligands, we needed another biophysical technique. STD NMR offered a robust method in determining whether these compounds were true binders. STD NMR confirmed pleurotin and purpurogallin as binders. There were solubility issues for many of the compounds, with visible precipitation when the compounds was dissolved in aqueous solution. This may be the reason for the poor quality of the NMR data which meant we could not say with confidence whether they were true ligands or not. It is however interesting to note the NSC215718 compound. It contains a thiourea group flanked by a halogenated benzene and hydroxyl naphthalene ring. It is structurally similar to the compound 'HS-152' (Figure 3.20). HS-152 is a compound recently discovered using a cell

based high throughput screening method by Tian et al as an inhibitor for Smurf 1 and Smurf2 which are part of the Nedd4 ubiquitin ligase family. The authors reported the compound may be an antagonist of the HECT catalytic capability. HS-152 is structurally similar as it contains a urea group flanked by two substituted benzene rings, one being halogenated and the other linked to a piperazine via a ketone. Analogues of HS-152 and NSC215718 could be synthesised and the presence of any SAR could be evaluated. A possible reason for NSC215718 not showing any inhibition is due to poor solubility.

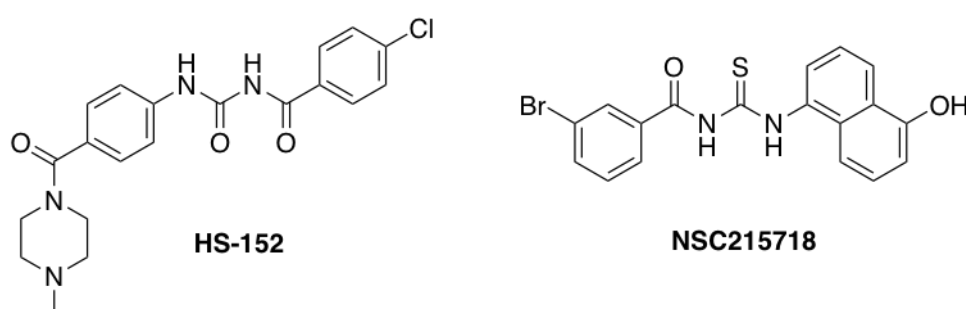


Fig. 3.20 -HS-152 and NSC215718

This chapter reported the first synthesis of heclin in only 3 steps using readily available starting materials. Beginning with a Vilsmeier-Haak reaction, **1** was converted to **2**. Reacting **2** in a Knoevenagel Doebner modification condensation reaction resulted in **4**. An amidation with **4** and 4-acetyl aniline yield the heclin molecule. This synthesis means large quantities of heclin can be taken forward for further testing. The nature of this synthesis lends itself to deriving analogues. Various substituted heterocycles could be used as a starting point for the formylation reaction. There is a multitude of anilines that could be coupled to the carboxylic acid, or simply any other cyclic amine could be used. The later part of this chapter described a brief attempt in NSC3567 analogue synthesis. NSC35676 came out as the most promising compound from the thermal shift assay as it was validated by STD NMR and SPR. The compound as been reported as an inhibitor for other proteins such as TLR1/2 (Cheng et al., 2012) and IP5 2-K (Whitfield et al., 2018). Therefore it is particularly important to derivatise the compound to make it specific to WWP2. Simply adding a methyl ether

group to the seven membered ring of the tropolone appeared to increase its activity. This was evident in both WWP2 and Nedd4. There is an argument that NSC3576 is a pan assay interference compound. The structure contains the catechol moiety (Figure 3.21), which has been implemented as a moiety which produced false positives (Jasial et al., 2017).

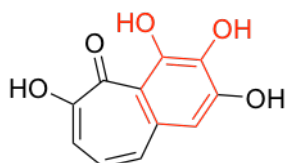


Fig. 3.21 Purpurogallin containing the catechol moiety

It is interesting when this moiety is eliminated in compound **11**, that all inhibition is lost. However the fact that NSC35676 was confirmed as a binder using thermal shift, NMR, SPR and has inhibitory activity makes a more convincing argument that it is a true inhibitor. Whilst it would be intriguing to carry on synthesising analogues of NSC35676, we felt the need for structural evidence to guide any future synthesis. This chapter set out to provide an experimental toolkit which can be used to identify WWP2 ligands and to also synthetically build upon the previously known inhibitor and a new inhibitor reported in this chapter. The next step in inhibitor development is to investigate inhibition/ binding mechanisms and to synthetically modify hit compounds. In order to do this, one must look at how the compounds are binding from a structural perspective.

Chapter 4

**Towards the solving of the first ligand
bound WWP2 crystallographic structure**

4.1 Introduction

Obtaining an X-ray structure of a ligand bound to a protein is a significant step in the development of small molecule inhibitors. It is considered paramount in the validation of hit compounds identified during high throughput screening. Structural information will unveil ligand binding sites and the orientation of the ligand. This information can also be used towards the understanding of any inhibition mechanisms. Interactions with the ligand and the surrounding amino acids can be deduced. Using organic synthesis, these interactions can then be exploited by developing the initial ligand into targeting specific residues, with the aim of developing a superior ligand. In order to gain a structure of a protein ligand complex, pure homogeneous protein must first be acquired. There are then two main approaches to introducing the ligand; soaking or co-crystallisation. Soaking involves growing apo protein crystals and soaking the crystals in a solution of chosen ligand. Co-crystallisation involves the growing of protein crystals in the presence of the ligand.

4.1.1 Crystal formation

Crystal formation encompasses two key steps; nucleation and growth. Nucleation is the process where protein molecules undergo the phase transition from the disordered aqueous phase to the ordered solid state (García-Ruiz, 2003). The exact process as to how this occurs is still the subject of some ambiguity (McPherson, 2004) (García-Ruiz, 2003) (Van Driessche et al., 2018), however it is presumed protein molecules aggregate into small ordered nuclei. Once nucleation has occurred, the much better characterised process of macromolecular crystal growth can start. The formation of what is known as a supersaturated state is key to both nucleation and growth. Supersaturation is when the solubility limit of the macromolecule has exceeded. Once supersaturation is reached, protein molecules still face a free energy barrier to reach the solid phase: crystal nucleus or precipitate. As supersaturation increases,

the energy barrier to nuclei formation decreases (Garcia-Ruiz, 2003). The nuclei will continue to grow into a crystal or precipitate as long as this supersaturation state is maintained. To achieve supersaturation, one is effectively decreasing the solubility of the macromolecule in the mother solution (McPherson, 2004). The protein in the mother solution, which is either at its solubility limit or below, is exposed to a precipitant solution. The composition of this precipitant solution promotes supersaturation. The precipitant solution can do this through a variety of ways such as changing the salt concentration, altering the pH and adding a particular precipitant to alter the interactions between the solute and solvent (McPherson, 2004).

4.1.2 X-ray diffraction

The result of the X-ray diffraction experiment is a map of electron density. Monochromatic X-rays are directed at the crystalline solid and scatter upon encountering electrons. The scattering is recorded to produce the diffraction pattern. The diffraction pattern is made up of numerous spots in different positions and of varying intensities. Measuring the intensities of each spot is used in the calculation of the electron density map via a Fourier Synthesis. However the phase shift, which is needed to solve the electron density distribution using the Fourier Synthesis, is unknown. This is known as the "phase problem" (Taylor, 2003). One of the simplest methods for solving the phase problem is to use molecular replacement. Molecular replacement uses a homologous structure (known as the search model) to orientate the protein within the unit cell. The phases from the search model can then be used with the experimentally measured amplitudes to generate the initial map (Evans and McCoy, 2008). Finally, the generated structure undergoes refinement. Refinement represents minimizing the difference between the experimental and calculated structure factors. The theoretical values of the physical properties of the polypeptide such as bond angles, bond length, phi and psi angles, dihedral angles planarity, clashes etc are taken into account when modifying

the model. This interpretation of the three dimensional electron density map combined with computational refinement leads to the final model of the target macromolecule.

4.1.3 Protein ligand structures of the HECT ubiquitin ligases

As of 2019, there is still only one published HECT ubiquitin ligase structure with a small molecule bound. As described in Section 1.2.3.3, this is the Nedd4 HECT bound to the indole based inhibitor (Kathman et al., 2015). The structure showed the compound covalently binding to a non catalytic cysteine through a Michael addition reaction. The presence of this bound inhibitor prevented the ubiquitin chains growing in a processive manner and changed it to a distributive manner. The structure not only provided a novel mechanistic insight into inhibiting Nedd4 HECT but also set a basis for further inhibitor development. (Quirit et al., 2017).

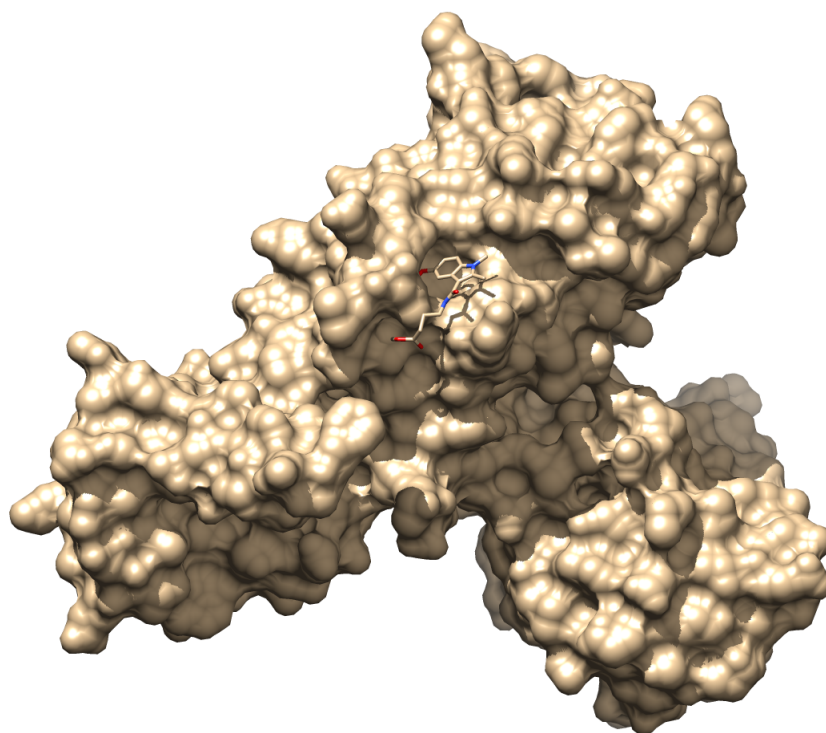


Fig. 4.1 -Nedd4 HECT bound to "Compound 1". "Compound 1" covalently binds to the non catalytic cysteine of the HECT domain, preventing the build up of ubiquitin chains. PDB: 5C91.

4.1.4 Chapter aims

Elucidating a protein ligand structure is a substantial milestone in rational inhibitor design. This chapter sets out to elucidate the first structure of an inhibitor bound to WWP2 by X-ray crystallography. Inhibitors and ligands described in Chapter 3 and also discovered in the Chantry lab (Watt et al., 2018) (Watt, 2018). are used in an attempt to generate a structure.

4.2 Results

4.2.1 Identification of putative binding sites

In silico docking was used to gain an idea as to where we might expect extra electron density to be picked up which could possibly be attributed to a ligand binding event. The WWP2 HECT (aa 486-865) crystal structure (PDB:4Y07) was used as the model for identifying putative binding sites. SiteMap software in the Schrodinger Maestro suite (Halgren, 2009) (Halgren, 2007) was used to search for binding sites. Two binding sites were identified based on pocket size large enough to accommodate a small molecule, solvent exposure, balance of region hydrophobicity and hydrophilicity and potential to form hydrogen bonds (Halgren, 2009). Site 1 (Figure 4.2, (A)) is located close to the hinge loop, sitting between the N and C lobe. The second potential binding site was located in the N-lobe (Figure 4.2 (B)). Neither sites have catalytic activity or known involvement in ubiquitin and E2 binding.

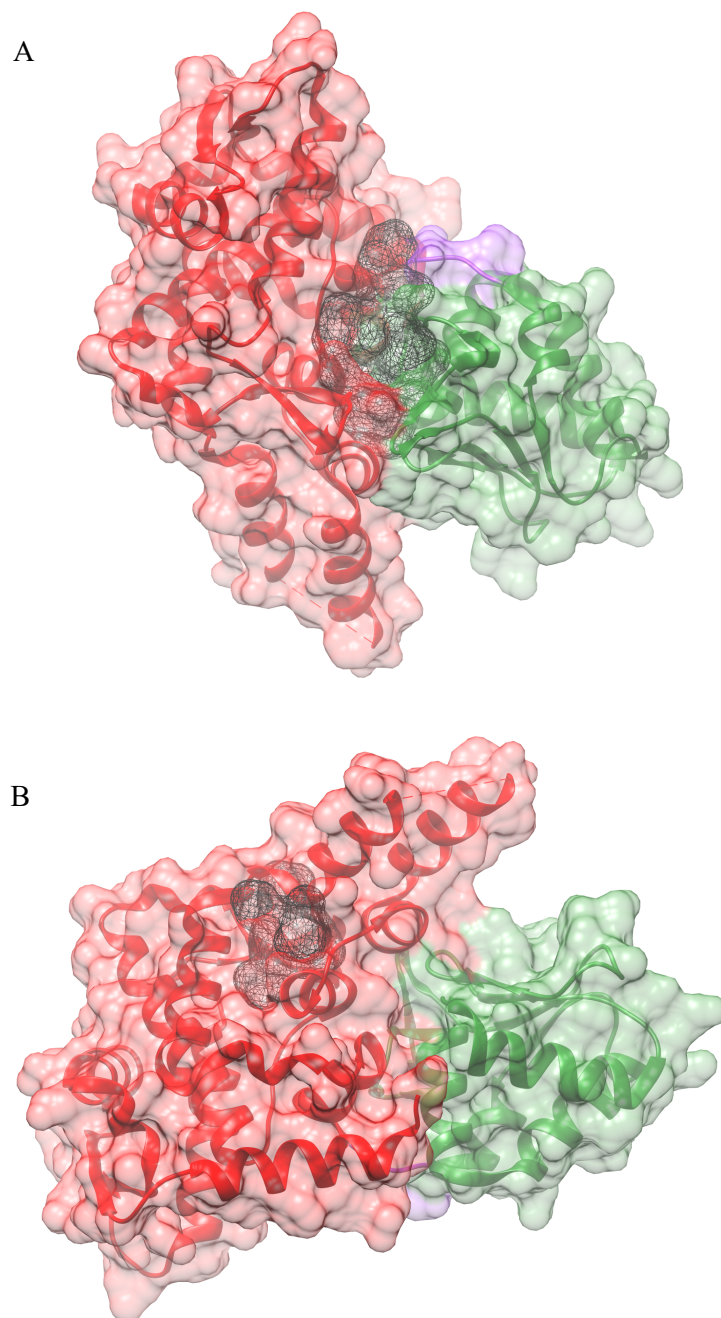


Fig. 4.2 Predicted small molecule binding sites for WWP2 HECT. (A) Site 1 is identified as a region sitting between the interface of the N and C lobe, in close proximity to the hinge loop. (B) Site 2 is identified as a region within the N lobe. The binding regions are shown in black mesh. The N lobe is coloured in red, the C-lobe in green and the hinge loop in purple.

4.2.2 Attempting WWP2 HECT crystallisation

Early efforts to crystallise WWP2 was based on the method described by Gong et al. (2015), described briefly in Section 3.2.1.1. Unlike the previous biophysical methods we tried with WWP2 HECT, where the His- tag could be left on, it was deemed the removal of the tag would most likely preclude crystallisation success. His- tagged WWP2 HECT was treated with thrombin and incubated overnight at 4 °C. Thrombin is most active at 16 °C, however the WWP2 HECT protein was unstable at this temperature and precipitated out of solution. A partial digest was observed, shown in Figure 4.3. In this not uncommon scenario, the heterogeneous mixture of tagged and untagged protein would be separated by running the sample down a Ni- NTA column. However, when this was tried with WWP2 HECT, untagged HECT has a strong affinity to Ni- NTA, preventing the isolation of the untagged variant. We tried incubating with thrombin for up to 70 hours in an attempt to get a full digest. We also tried dialysing the protein back into PBS and incubating with thrombin for a further 3 days, as PBS is the standard buffer for thrombin. Eventually a 80-90 % digest was observed and the protein was further purified by gel filtration (Figure 4.3) and concentrated to 10 mg mL⁻¹. Sitting drop vapour diffusion plates were set up using a small screen around the original precipitant conditions. 48 conditions using 0.1 M HEPES, pH 8.2-8.5, 0.2 M MgCl₂ and 12-17 % ethanol were trialled. No crystals were observed over the following weeks. We attributed the lack of success to be most likely due to not yielding a fully homogenous sample. Conveniently, in 2017, Z.Chen et al published another crystal form of WWP2 while investigating the role of the WW2 domain and the linker regions mediating HECT activity. Two constructs were used: WW2- 2,3 linker- HECT and 2,3 linker- HECT produced crystals. The WW2- 2,3 linker-HECT crystals were reported to give a higher resolution than the 2,3 linker- HECT crystals. We chose to leave the HECT (886-865 aa) construct and choose the WW2- 2,3 linker HECT construct in pursuit of WWP2 crystals.

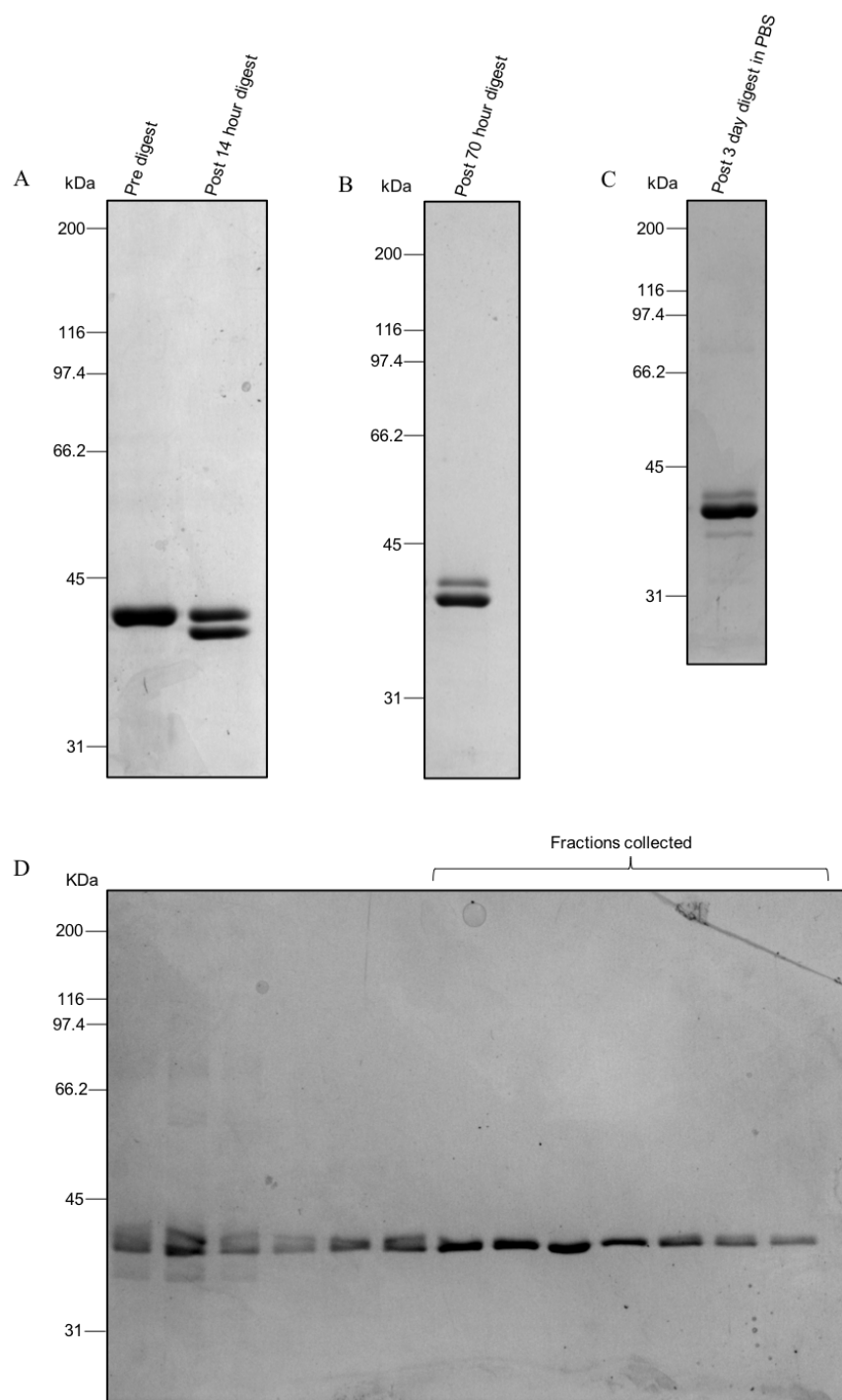


Fig. 4.3 -WWP2 HECT purification for crystallography. (A) Incubating WWP2 HECT with thrombin overnight resulted in approximately a 50 % digest. (B) Incubating WWP2 HECT for a further 2-3 days yielded a greater portion of His tag free WWP2 HECT. (C) WWP2 HECT after being dialysed into PBS and incubated for a further 3 days with thrombin. Only slightly more His tagged protein was digested. (D) WWP2 gel filtration SDS PAGE gel. The fractions which visibly looked the most homogenous were collected and concentrated for the crystallography experiments.

4.2.3 WWP2 WW2-2,3 linker HECT purification and crystallisation

The pGEX6p-2 WWP2 WW2-2,3 linker HECT (WW2-L-H) plasmid was gifted by Professor Philip Cole. The method described in Chen et al. (2017) was followed with some minor changes. The protein was expressed in CodonPlus RP as a glutathione s-transferase (GST) fusion protein and purified using a GS Trap column. Cleavage of GST was carried out using PreScission Protease overnight at 4 °C. The sample was further purified using size exclusion chromatography. Any remaining GST could not be separated during gel filtration due to the small difference between GST and the WWP2 protein. GST is 26 kDa in size but exists as a dimer, which brings it to a similar size to the WW2-L-H construct, rendering the gel filtration column unable to resolve the two proteins. Remaining GST could be removed by a final wash through the GS trap. The protein was initially concentrated at 4 °C using vivaspin ultracentrifugation tubes. However, there was an increasingly dramatic loss in protein when concentrating above 1 mg mL⁻¹. We noted a temperature driven effect with the solubility of the protein. Pure protein sample at 4 °C would begin to precipitate out, upon heating to room temperature the precipitate would dissolve back into solution. Hence, concentrating at 20 °C minimised loss of protein and we concentrated the sample to 1-7 mg mL⁻¹ for crystallisation trials. Various stages of the purification process are shown by SDS-PAGE in Figure 4.4. Conditions in the sitting drop method were set up around the two published crystallisation conditions of 0.1M KSCN, 30 % w/v PEG monomethyl ether 2000 (‘condition 1’) and 0.1 M MMT (malic acid, MES and tris in molar ratio of 1:2:2) pH 6.0, 25 % w/v PEG 1500 (‘condition 2’). Initial efforts produced needle clusters in condition 1 and micro crystals in condition 2 at 3 mg mL⁻¹. Upon reaching the higher concentration of 7 mg mL⁻¹, prismatic crystals formed in condition 2 after 1 week.

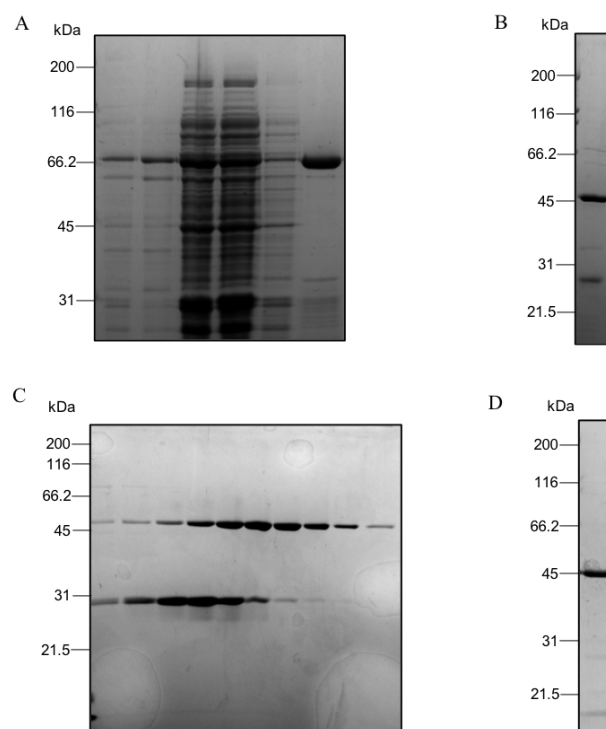


Fig. 4.4 -Key steps in the WW2-2,3 linker HECT purification. (A) GS trap purification. (B) GST digest with PreScission protease. Note the undigested protein at 66.2 kDa and the GST at 29 kDa. (C) Gel filtration of the protein. (D) Final purified protein for crystallography.

4.2.4 Initial WWP2 crystal harvesting

A total of 28 suitable crystals were picked during the first harvest. 15 crystals were set aside for apo structure determination and 13 crystals were soaked in three different ligands. All ligand stocks were dissolved in DMSO- d_6 at 50 mM. Crystals were soaked in NSC2805, NSC288387 and NSC401005 at 10 x their approximate IC₅₀ concentrations at 0.05 % DMSO. Crystals were cryoprotected in the soak solution with added 30 % glycerol. Data was collected at the Diamond Light Source (Oxford). The data was processed and integrated using the automatic pipelines available on ISPyB. Molecular replacement was carried out using the PHENIX suite and rounds of automatic and manual refinement was done using COOT. 7 crystals gave suitable data for structural analysis. Initial molecular replacement was done using the original WWP2 WW2 2,3 linker HECT structure (PDB:5TJ8, NB: no WW2

domain observed). All the solved structures matched in high accordance with the search model. Resolutions ranged from 2.05-3.58 Å, with the best data set having better resolution than the original search model. Close inspection of the 2Fo-Fc electron density maps did not yield any significant unexplained density.

Crystal code	Experiment type	Ligand	Ligand concentration μM	DMSO concentration %
WWP2_1_1	Apo	N/A	N/A	N/A
WWP2_4_1	Apo	N/A	N/A	N/A
WWP2_5_2	Soak	NSC288387	11.5	0.05
WWP2_5_4	Soak	NSC2805	2	0.05
WWP2_6_1	Soak	NSC2805	2	0.05
WWP2_6_3	Soak	NSC2805	2	0.05
WWP2_7_2	Soak	NSC401005	13	0.05

Table 4.1 -Summary of first soaking experiments. Information regarding the harvested crystals for which data was not collected is not shown.

Crystal code	Pipeline	Resolution Å
WWP2_1_1_2	xia2 3dii	2.74
WWP2_4_1_2	xia2 dials	2.44
WWP2_5_2_2	fast_dp	2.62
WWP2_5_4_2	xia2 dials	2.48
WWP2_6_1_3	autoProc	2.05
WWP2_6_3_2	autoProc	3.58
WWP2_7_2_2	xia2 dials	2.8

Table 4.2 -Summary of the solved data from the first soaking experiments.

4.2.5 Further soaking experiments leads to improved search model

Further soaking experiments were conducted with higher concentrations of the inhibitors. Reaching higher concentrations of inhibitor meant pushing the DMSO percentage higher. Concentrations of 5 mM compound, 10 % DMSO and 2.5 mM compound, 5 % DMSO were attempted. A greater range of inhibitors were chosen; NSC288387, NSC2805, NSC401005, NSC35676, Heclin and Aquaymycin. Aquaymycin is a natural product which was discovered as a WWP2 inhibitor in the Chantry lab (Watt, 2018) whilst this research was being carried out. Only one crystal subjected to 10 % DMSO gave Bragg diffraction, and at the low resolution of 5 Å. Initial soaks where the crystal was subjected to 5 % DMSO gave high resolution structures. This suggested the crystals could withstand this DMSO concentration, therefore setting the precedent for the remaining soaking experiments outlined in Table 4.3. Crystals which gave bragg diffraction were subjected to data processing, integration, molecular replacement and refinement as outlined in Section 4.2.4. WWP2_13_2 gave the highest resolution of 1.99 Å, however no ligand was observed in the electron density map. As this data set was of a higher quality than previously published, the model was refined through reiterate rounds of manual and automatic refinement in Coot in order to be used as an improved search model for molecular replacement.

Crystal code	Experiment type	Ligand	Ligand concentration mM	DMSO concentration %
WWP2_13_1	Soak	NSC288387	2.5	5
WWP2_13_2	Soak	NSC288387	2.5	5
WWP2_14_2	Soak	NSC288387	5	10
WWP2_15_4	Soak	NSC2805	2.5	5
WWP2_17_1	Soak	NSC401005	2.5	5
WWP2_17_2	Soak	NSC401005	2.5	5
WWP2_17_3	Soak	NSC401005	2.5	5
WWP2_18_3	Soak	NSC2805	2.5	5
WWP2_18_4	Soak	NSC2805	2.5	5
WWP2_19_1	Soak	NSC2805	2.5	5
WWP2_19_2	Soak	NSC2805	2.5	5
WWP2_19_3	Soak	Heclin	2.5	5
WWP2_19_4	Soak	Heclin	2.5	5
WWP2_21_2	Soak	Aquaymycin	2.5	5
WWP2_22_2	Soak	NSC288387	2.5	5
WWP2_22_3	Soak	NSC288387	2.5	5
WWP2_22_4	Soak	NSC288387	2.5	5
WWP2_23_1	Soak	NSC288387	2.5	5
WWP2_23_3	Soak	NSC401005	2.5	5
WWP2_24_2	Soak	NSC401005	2.5	5
WWP2_24_3	Soak	NSC401005	2.5	5
WWP2_24_4	Soak	NSC401005	2.5	5
WWP2_26_1	Soak	NSC35676	2.5	5
WWP2_26_2	Soak	NSC35676	2.5	5
WWP2_26_3	Soak	NSC35676	2.5	5
WWP2_26_4	Soak	NSC35676	2.5	5
WWP2_27_1	Soak	NSC35676	2.5	5
WWP2_27_2	Soak	NSC35676	2.5	5

Table 4.3 -Summary of second soaking experiments

Crystal code	Pipeline	Resolution Å
WWP2_13_1	xia2 3dii	2.21
WWP2_13_2	xia2dials	1.99
WWP2_14_2	autoProc	5.04
WWP2_15_4	xia2 3dii	4.69
WWP2_17_1	xia2 3dii	3.33
WWP2_17_2	xia2 3dii	3.09
WWP2_17_3	autoProc + STARANISO	2.31
WWP2_18_3	xia2 3dii	4.27
WWP2_18_4	fast_dp	3.12
WWP2_19_1	autoProc	3.53
WWP2_19_2	autoProc	4.56
WWP2_19_3	xia2 3dii	3.4
WWP2_19_4	xia2 3dii	3.64
WWP2_21_2	xia2 3dii	3.38
WWP2_21_3	xia2 3dii	4.08
WWP2_21_4	xia2 3dii	5.02
WWP2_22_2	autoProc	2.63
WWP2_22_3	xia2 3dii	3.35
WWP2_22_4	autoProc	3.57
WWP2_23_1	autoProc + STARANISO	2.4
WWP2_23_3	autoProc + STARANISO	2.68
WWP2_24_2	autoProc + STARANISO	2.44
WWP2_24_3	xia2dials	3.04
WWP2_24_4	xia23dii	3.42
WWP2_26_1	xia2dials	2.7
WWP2_26_2	xia2dials	3.01
WWP2_26_3	autoProc + STARANISO	3.08
WWP2_26_4	autoProc	2.82
WWP2_27_1	autoProc + STARANISO	2.47
WWP2_27_2	autoProc + STARANISO	2.51

Table 4.4 -Summary of the solved data from the second soaking experiments

	wwp2_13_2	Published 5tj8
Data collection		
Wavelength Å	0.9795	1.0000
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Cell constraints a, b, c	44.23, 91.37, 112.22	44.23, 89.27, 111.47
α, β, γ	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Resolution	56.11-1.99 (2.02-1.99)	50-2.3
% Data completeness (in resolution range)	99.9 (99.6)	95.5
Number of unique reflections	32006 (1539)	
Multiplicity	12.9 (13.3)	
Rmerge	0.024 (0.531)	0.08
Rmeas	0.116 (2.663)	
$\langle I/\sigma(I) \rangle$	12.8 (1.0)	3.02
Wilson B factor	41	
Refinement		
Resolution range	56.11-1.99 (2.02-1.99)	
Completeness	99.9 (99.6)	
Reflections used in refinement	31974 (3099)	
Reflections used for R _{free}	1536	
R _{work}	0.206	0.215
R _{free}	0.251	0.316
Number of non hydrogen atoms	3560	
Protein residues	415	
RMS(bonds)	0.007	
RMS(angles)	0.841	
Ramachandran favoured	91.93	92
Ramachandran outliers	1.71	2.3
Rotamer outliers	0	0
Average B factor	70.0	67.0

Table 4.5 -Validation data comparison of the new search model and the published structure. Data in brackets represents the high resolution bin.

4.2.6 First potential protein ligand structure

Upon re doing molecular replacement and refinement of the data collected outlined in Table 4.2 with the new search model, an area of interesting unexplained electron density was picked up in the electron density map of WWP2_6_1. Interestingly the area of additional density was identified in the *in silico* predicted binding site 1. The corresponding ligand used for the

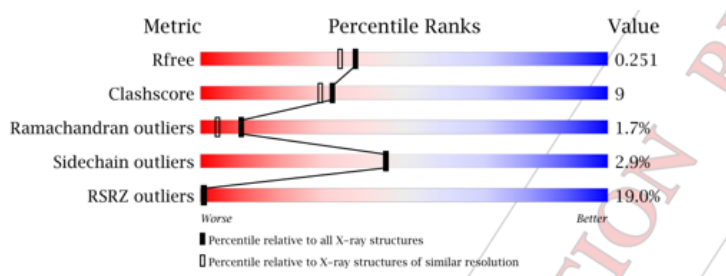
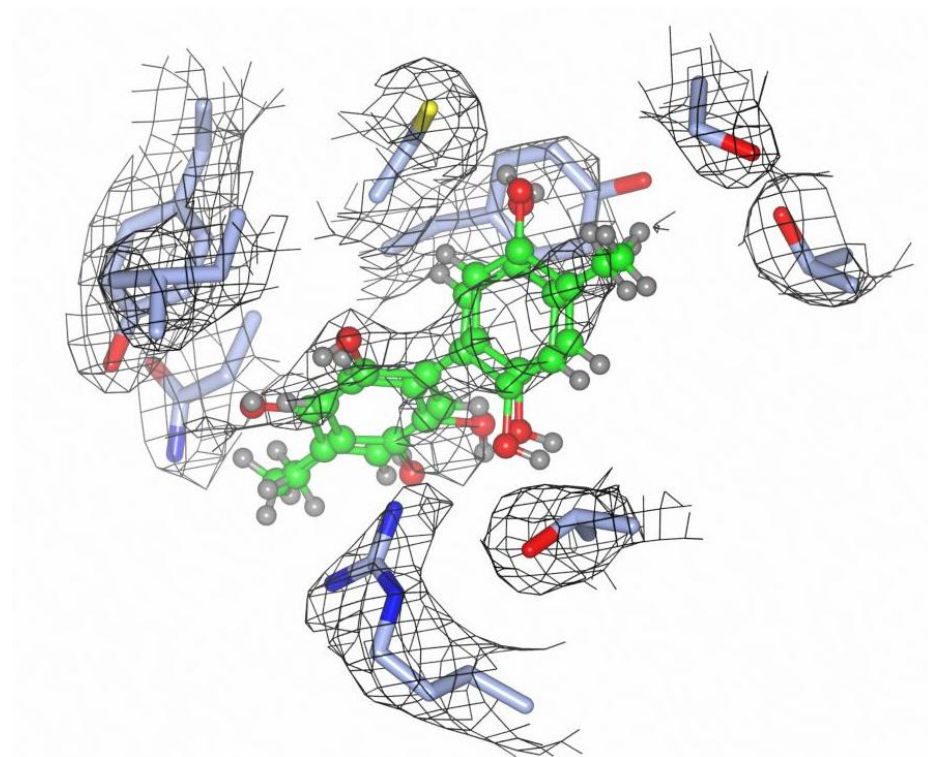


Fig. 4.5 -PDB validation of new search model

WWP2_6_1 soak was the biphenyl tetrol compound NSC2805. Restraints for the ligand was built in eLBOW and the ligand was fitted into the unexplained density. Owing to the available axial rotation in the symmetrical biphenyl, one of the rings was modelled to have 2 rotomers with differential occupancy. The electron density maps of the ligand are shown in Figure 4.6. The ligand is surrounded by Thr801, Cys802, Arg803, Tyr587, Cys588, Asn586, Leu589 and Thr837. π - π interactions can be observed with Tyr587 at a distance of 4.1 Å. Arg803 forms a π - cation interaction. Putative hydrogen bond interactions are present between the hydroxyls of NSC2805 and Tyr587, Thr801, Cys 588, Leu589, Tyr581 and Arg803. However, at a distance of 2.88- 3.64 Å these interactions would be weak as their bond lengths stretch to the limits of conventional hydrogen bonds. The protein ligand interactions are shown in Figure 4.7.

A



B

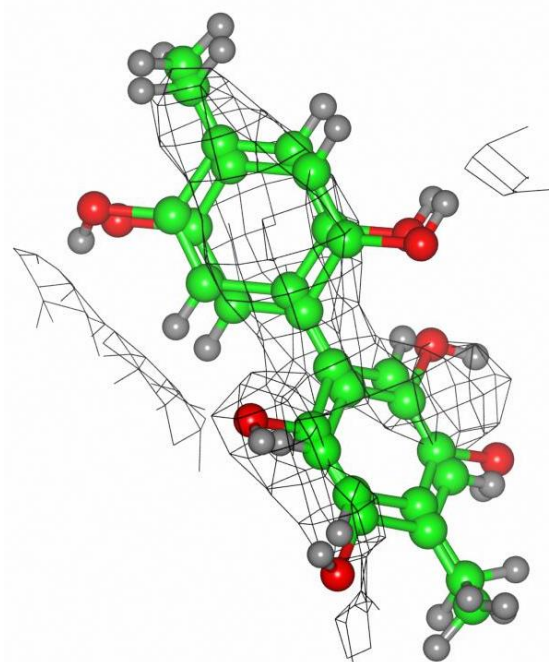
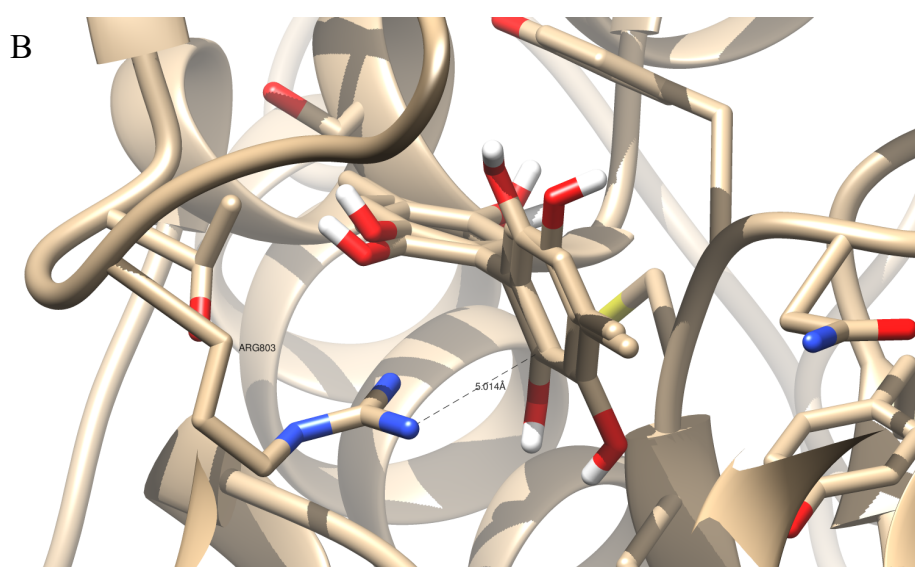
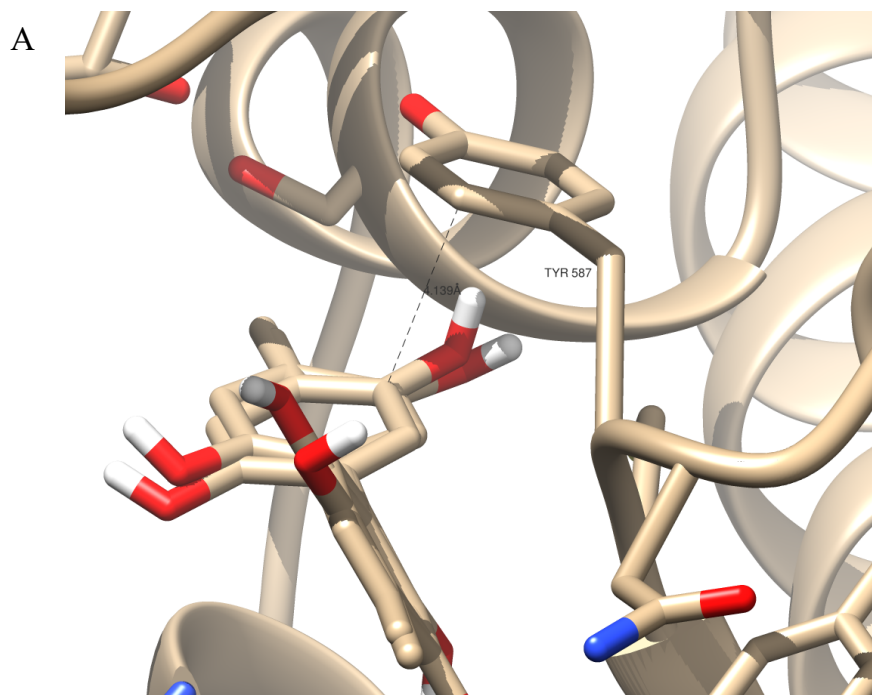


Fig. 4.6 -WWP2_6_1 double difference Fourier electron density map. (A) Electron density of the ligand and surrounding residues (B) Electron density of the ligand. The rotomer at 0.45 occupancy has a RSCC of 0.61 and the rotomer at 0.55 occupancy has a RSCC of 0.68. Map contoured at 0.9 σ .



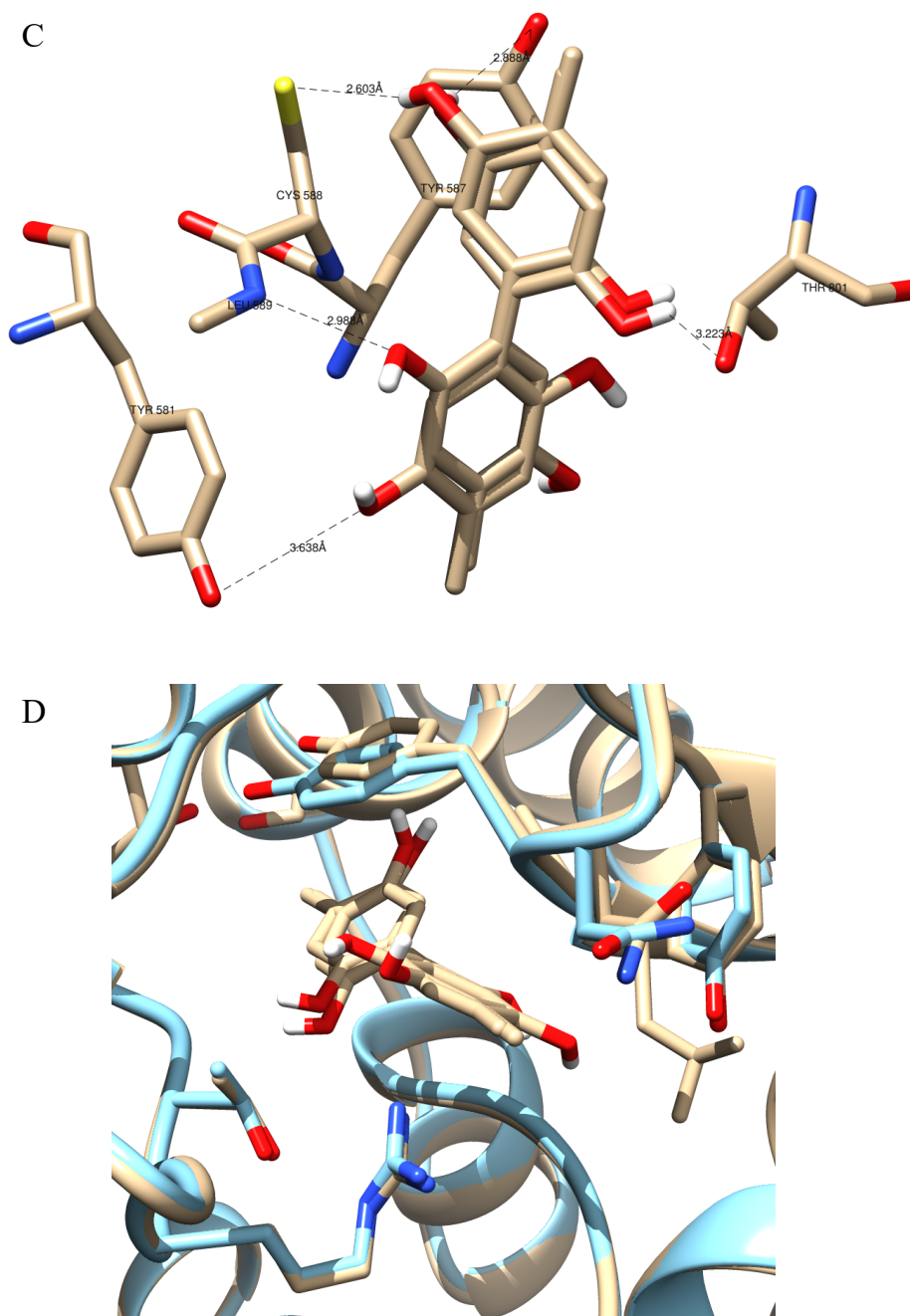


Fig. 4.7 -Protein ligand interactions present in WWP2_6_1_3. (A) π - π stacking between Tyr587 and NSC2805. (B) π - cation interaction with Arg803. (C) Possible hydrogen bonds between the ligand and the surrounding amino acids. (D) Structure comparison of the apo (blue) and ligand bound (brown) model.

Data Collection	
Wavelength Å	0.9795
Space group	P2 ₁ 2 ₁ 2 ₁
Cell constraints a, b, c	44.03, 89.73, 111.27
α, β, γ ,	90.00, 90.00, 90.00
Resolution	44.86- 2.05 (2.13-2.05)
% Data completeness(in resolution range)	98.75 (89.24)
Total reflections	159627 (12942)
Unique reflections	28278 (2502)
Multiplicity	5.6 (4.7)
R _{merge}	0.1006 (1.846)
R _{meas}	0.111(2.079)
$\langle I/\sigma(I) \rangle$	10.86 (0.79)
Wilson B factor	41.67
Refinement	
Resolution range	44.86- 2.05 (2.13-2.05)
Completeness	98.75 (89.24)
Reflections used in refinement	27980 (2497)
Reflections used for R _{free}	1406 (137)
R _{work}	0.23 (0.40)
R _{free}	0.30 (0.45)
Number of non hydrogen atoms	3553
Protein residues	415
RMS(bonds)	0.008
RMS(angles)	1.19
Ramachandran favoured	91.44
Ramachandran outliers	2.2
Rotamer outliers	2.94
Average B factor	68.95

Table 4.6 -WWP2_6_1 validation. Data in brackets represents the high resolution bin.

4.2.7 Optimising of soak experiments to reveal further protein ligand structures

Optimisation of the protein purification by simply shortening the time from expression to crystal plate set up resulted in a greater number of crystals (2-4 crystals per plate to 100s per plate). Owing to the availability a greater number of crystals, we investigated further as to whether high number of sub- standard data sets was due to either ligand, cryoprotection or soaking condition effects. Apo crystals were cryoprotected in 10 %, 20 % and 30 % glycerol (summarised in Table 4.7).

Crystal code	Experiment type	Glycerol content present in cryoprotectant %
WWP2_27_3	Apo	0
WWP2_27_4	Apo	30
WWP2_28_1	Apo	10
WWP2_28_2	Apo	20

Table 4.7 -Summary of glycerol titration soaking experiments

Crystal code	Pipeline	Resolution Å
WWP2_27_3	autoProc	2.06
WWP2_27_4	xia2dials	2.13
WWP2_28_1	autoProc	2.15
WWP2_28_2	xia2dials	1.96

Table 4.8 -Summary data collected from glycerol titrations

Crystal code	Experiment type	DMSO content present in cryoprotectant %
WWP2_31_1	Apo	1
WWP2_32_2	Apo	2
WWP2_32_3	Apo	2
WWP2_33_2	Apo	3
WWP2_33_3	Apo	3
WWP2_33_4	Apo	3

Table 4.9 -Summary of the DMSO soaking experiments

Crystal code	Pipeline	Resolution Å
WWP2_31_1	Xia2dials	2
WWP2_32_2	Xia2dials	2.1
WWP2_32_3	autoProc	2.48
WWP2_33_3	autoProc	2.23

Table 4.10 -Summary of the solved data from the DMSO soaking experiments

No trend in data quality was observed when altering the glycerol content in the cryoprotectant of apo WWP2 crystals (Table 4.8). 20 % Glycerol nominally gave the best resolution. All subsequent harvested crystals were cryoprotected with 20 % glycerol. Next we investigated whether the DMSO content was affecting the quality of the crystal. DMSO titrations in the cryoprotectant at concentrations of 1 %, 2 %, 3 %, and 5 % were carried out (summarised in Table 4.9). Up to 3 % DMSO, there was no deterioration in data quality (Table 4.10). However, visible cracking in the crystals was observed over time upon being subjected to the 3 % DMSO. We chose to pick 2.5 % DMSO concentration for all future soaking experiments, as we hypothesised this was the maximum DMSO concentration the crystals can handle without visible deterioration and loss of data quality, whilst being exposed to the highest concentration of ligand possible. It should also be noted various crystals forms were being observed at this point. Four main forms were observed; a square, 'extended square', hexagonal rod and large irregular blocks, there was no correlation in crystal form, precipitant solution and data quality.

The new soaking and cryoprotection conditions were used in attempting to improve the protein ligand structure reported in Section 4.2.6 (Table 4.11 and Table 12).

Crystal code	Experiment type	Ligand	Ligand concentration mM	DMSO concentration %
WWP2_34_2	Soak	2805	1.25	2.5
WWP2_34_3	Soak	2805	1.25	2.5
WWP2_34_4	Soak	2805	1.25	2.5
WWP2_35_1	Soak	2805	1.25	2.5

Table 4.11 -Summary of the NSC2805 soaking experiments in the improved soak conditions

Crystal code	Pipeline	Resolution Å
WWP2_34_2	autoProc	2.51
WWP2_34_3	Xia2dials	2.00
WWP2_34_4	autoProc	2.06
WWP2_35_1	Xia2dials	2.14

Table 4.12 -Summary of the solved data from the improved soaking experiments

The WWP2_34_3 data set presented two areas of additional unexplained electron density (Figure 4.8). The ligand was fitted into the density and the models were refined through iterative rounds of refinement. The ligand was bound to the original site reported in Section 4.2.6, albeit with a lower RSCC of 0.61 and an occupancy of 0.83. Unlike the structure in Section 4.2.6, a single rotamer was fitted. The ligand was present in the same space but was found to have a different orientation. The difference in orientation favoured the intermolecular protein ligand interactions. There are hydrogen bonds with Tyr587, Arg803 and three water molecules. The distances of these hydrogen bonds lie in the more typical range of 2.5-3.0 Å. There is the π - π stacking interaction with Tyr587 and the π -cation interaction with Arg803. Again, these distances are closer than in the WWP2_6_3 model at 3.8 Å and 3.7 Å. The new site was located on the outer surface of the N-lobe, close to an undefined region of the polypeptide chain. The density fit in this second binding site is much better with a RSCC of 0.75 with an occupancy of 1.0. The closest residues are Ser395 and Gly486. There is potential for hydrogen bonds with Ser395 and Gly486. The most likely interaction being the 2.4 Å hydrogen bond with the nitrogen on Gly486. Figure 4.9 shows the hydrogen bonds present in the two sites.

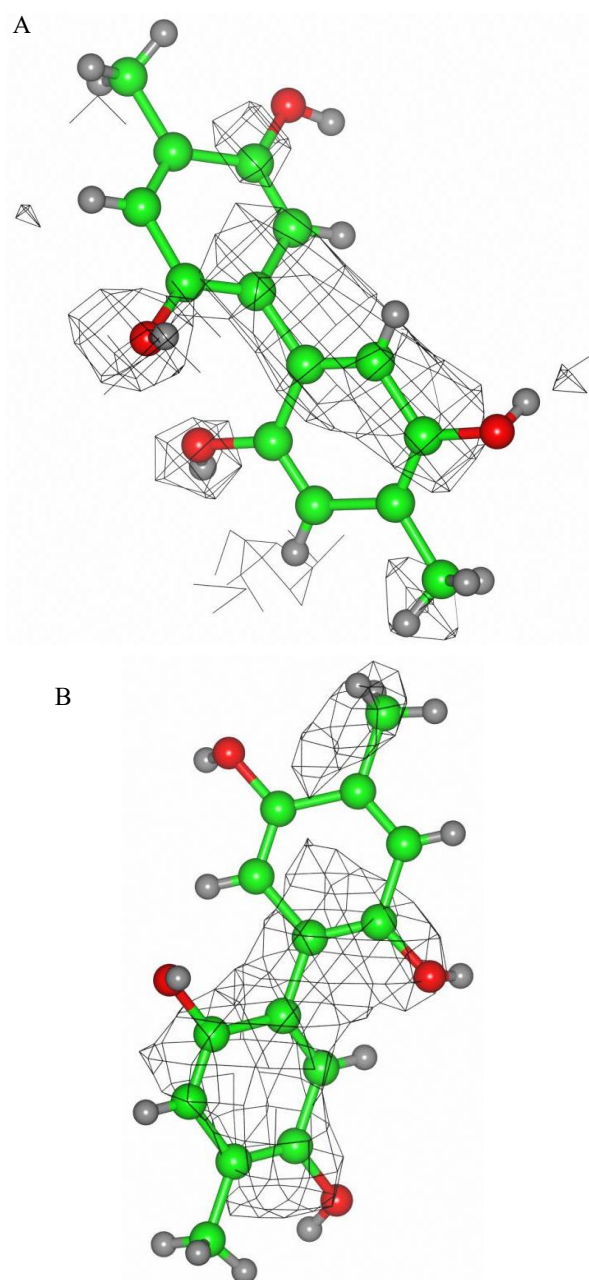
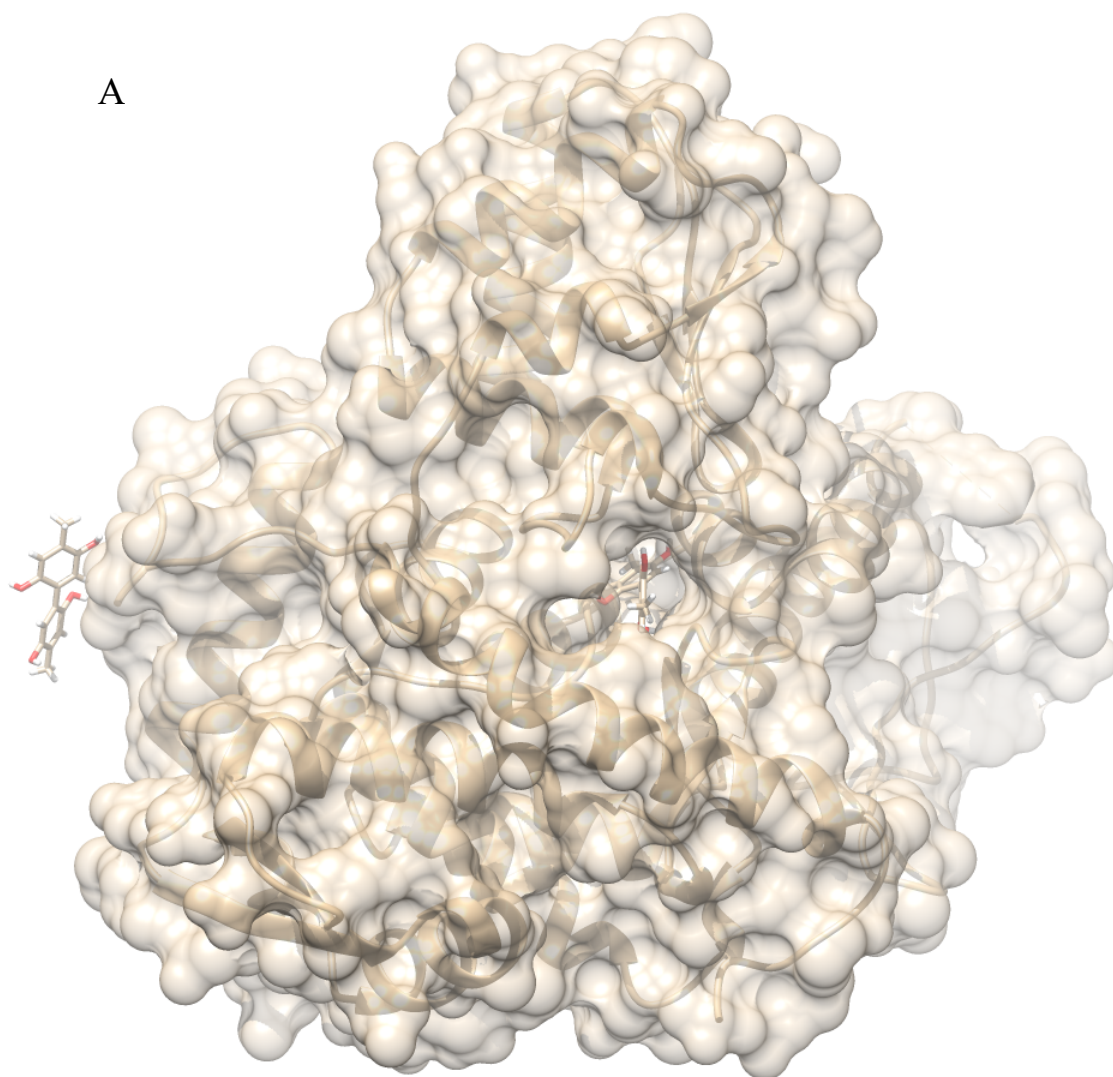


Fig. 4.8 -Double difference Fourier electron density of NSC2805 in WWP2_34_3. (A) Electron density of NSC2805 in the original site. The ligand has 0.86 occupancy and a RSCC value of 0.61. (B) Electron density of NSC2805 in the second identified site. The ligand has an occupancy of 1 and a RSCC value of 0.75. Map contoured at 0.9σ .



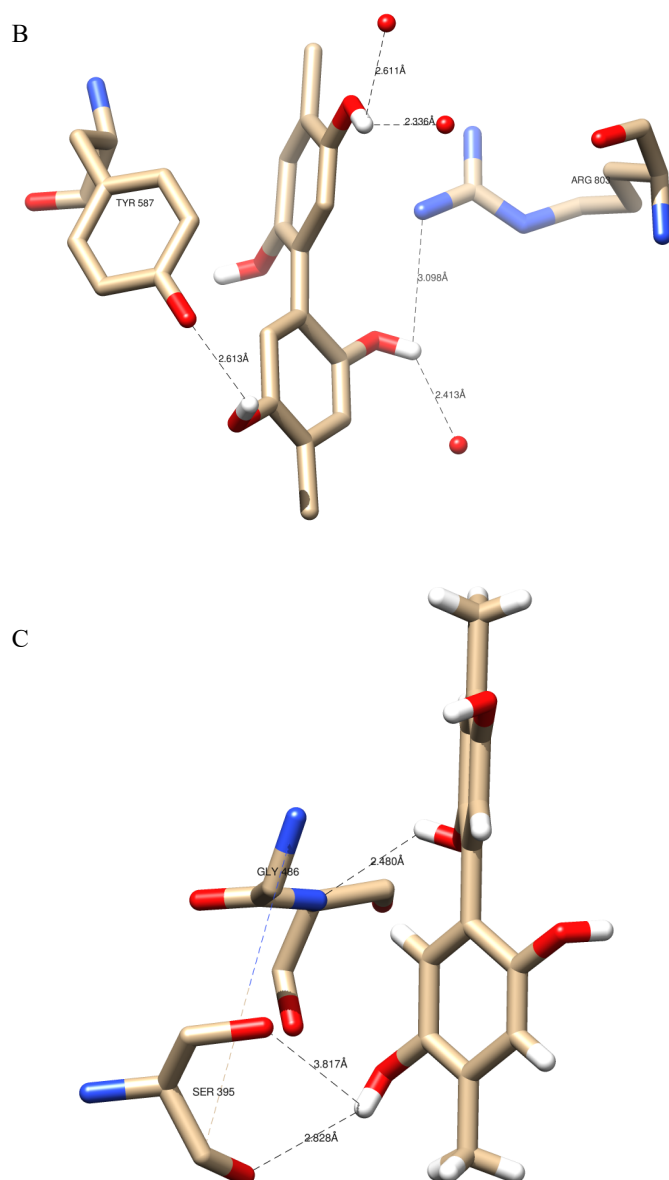


Fig. 4.9 -WWP2_34_3 protein ligand interactions. (A) Structure of WWP2 HECT with NSC2805 bound in the two sites. (B) Hydrogen bonds present in the original site. There are hydrogen bonds between 3 three waters at 2.61 Å, 2.34 Å and 2.41 Å. There is a 3.1 Å hydrogen bond with Arg803 and a 2.61 Å hydrogen bond with Tyr587. (C) Hydrogen bonds present in the second site. Gly486 makes a 2.48 Å hydrogen bond and Ser395 makes 2.83 Å and 3.82 Å hydrogen bonds.

Data Collection	
Wavelength A	0.9763
Space group	P2 ₁ 2 ₁ 2 ₁
Cell constraints a, b, c	44.11, 90.45, 111.93
$\alpha, \beta, \gamma,$	90.00, 90.00, 90.00
Resolution	55.97- 2.0 (2.07- 2.0)
% Data completeness (in resolution range)	92.33 (62.96)
Total reflections	305911 (10030)
Unique reflections	28674 (1910)
Multiplicity	10.7 (5.3)
R _{merge}	0.093 (1.232)
R _{meas}	0.097 (1.355)
$\langle I/\sigma(I) \rangle$	13.89 (1.06)
Wilson B factor	36.17
Refinement	
Resolution range	55.97- 2.0 (2.07- 2.0)
Completeness	92.33 (62.96)
Reflections used in refinement	28663 (1909)
Reflections used for R _{free}	1422 (79)
R _{work}	0.21
R _{free}	0.25
Number of non hydrogen atoms	3614
Protein residues	415
RMS(bonds)	0.007
RMS(angles)	1.13
Ramachandran favoured	90.95
Ramachandran outliers	1.47
Rotamer outliers	4.01
Average B factor	60.52

Table 4.13 -WWP2_34_3 validation. Data in brackets represents high resolution bin.

The resulting structures still left us with a desire to improve the resolution around the ligand to get rid of any ambiguity. All previous soaks had the ligand incubating with the crystal for a relatively short time, no longer than 1 hour. Therefore, we tried increasing the length of the ligand soak for 3 hours in our idealised soak conditions. Soaks with 1 % DMSO was also carried out in the event of the crystal being unable to tolerate the higher concentration of DMSO for a long period of time. The longer soaking experiments are characterised in Table 4.14 and data collected summarised in Table 4.15. The WWP2_38_1 data set showed the ligand in the first binding site, there was no improvement in the quality of electron density around the ligand (Figure 4.10).

Crystal code	Experiment type	Ligand	Ligand concentration mM	DMSO concentration %
WWP2_37_4	Soak	2805	0.5	1
WWP2_38_1	Soak	2805	0.5	1
WWP2_38_2	Soak	2805	0.5	1
WWP2_38_4	Soak	2805	0.5	1
WWP2_39_1	Soak	2805	1.25	2.5
WWP2_39_3	Soak	2805	1.25	2.5

Table 4.14 -Summary of long soak experiments

Crystal code	Pipeline	Resolution
WWP2_37_4	Xia2dials	3.02
WWP2_38_1	Xia2dials	1.98
WWP2_38_2	Xia2dials	2.32
WWP2_38_4	Xia2dials	2.22
WWP2_39_1	Xia2dials	2.55
WWP2_39_3	AutoProc + STARANISO	2.39

Table 4.15 -Summary of the solved data from the long soaking experiments

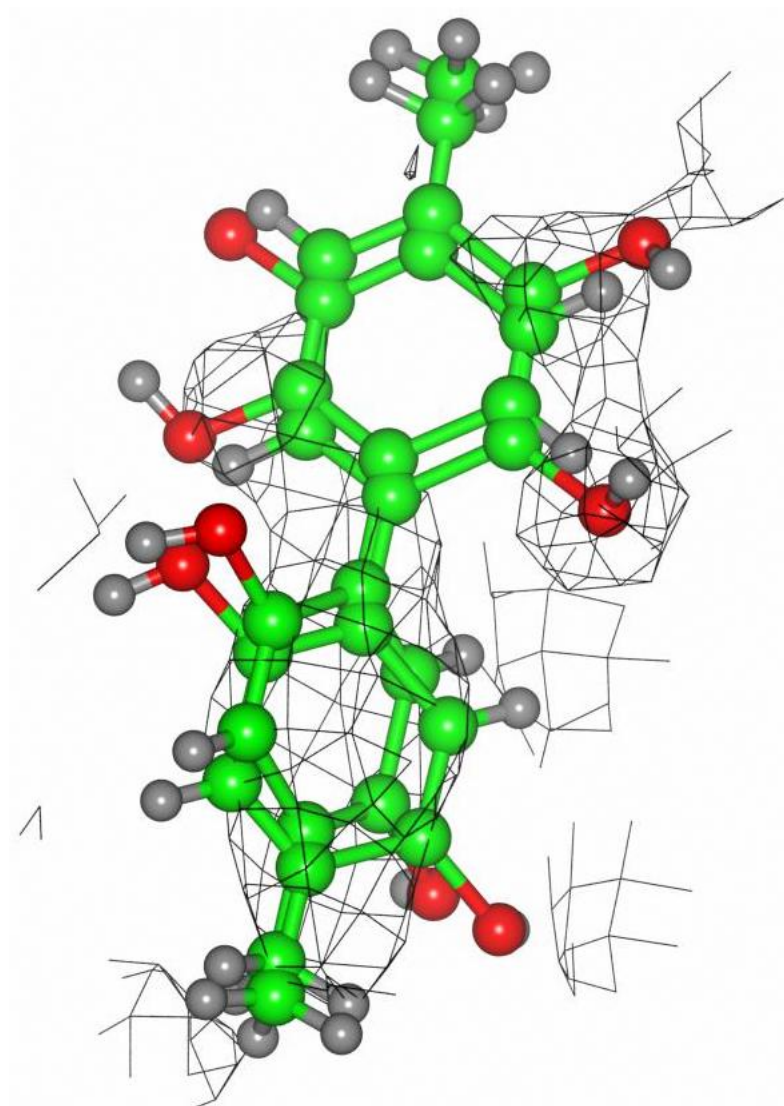


Fig. 4.10 -NSC2805 electron density in WWP2_38_1. One rotomer has an occupancy of 0.41 and a RSCC of 0.62 and the other rotomer has an occupancy of 0.59 and a RSCC of 0.67. Map contoured at 0.9σ .

Data Collection	
Wavelength A	0.9795
Space group	P2 ₁ 2 ₁ 2 ₁
Cell constraints a, b, c	44.09, 91.22, 112.02
$\alpha, \beta, \gamma,$	90.00, 90.00, 90.00
Resolution	47.73- 1.98 (2.05- 1.98)
% Data completeness (in resolution range)	99.85 (99.55)
Total reflections	423628 (39665)
Unique reflections	32253 (3135)
Multiplicity	13.1 (12.7)
R _{merge}	0.111 (1.784)
R _{meas}	0.1159 (1.859)
$\langle I/\sigma(I) \rangle$	11.68 (1.12)
Wilson B factor	40.35
Refinement	
Resolution range	47.73- 1.98 (2.05- 1.98)
Completeness	99.85 (99.55)
Reflections used in refinement	32216 (3129)
Reflections used for R free	1607 (164)
R _{work}	0.21 (0.30)
R _{free}	0.26 (0.34)
Number of non hydrogen atoms	3627
Protein residues	415
RMS(bonds)	0.008
RMS(angles)	1.15
Ramachandran favoured	91.20
Ramachandran outliers	1.71
Rotamer outliers	2.14
Average B factor	64.63

Table 4.16 -WWP2_38_1 validation table. Data in brackets represents the high resolution bin.

4.3 Discussion

Structure based drug design is a prominent method in drug discovery. The structure of a target protein bound to a small molecule is used to direct the design of subsequent compounds with the aim of delivering a finalised drug compound into clinical trials. Analysis of the structure can determine what theoretical molecular entities would increase or decrease favourable properties such as affinity and specificity. The structure data can be used in not only the hit to lead stage of drug design, but in lead optimisation. If the compound is found to have undesirable properties such as low aqueous solubility or a high metabolic turnover, alterations to the compound can be guided alongside the structure assuring the improved compound would still accommodate the binding pocket. Generation of protein ligand structures can also unveil the mechanism behind inhibition. Knowledge of inhibition mechanisms can in turn lead to new or improved types of inhibitors. The primary method for gaining a protein ligand structure is through X- ray diffraction. This involves growing suitable protein crystals which give high resolution diffraction. Compounds can be added before the growing of the crystal or just before the crystal is harvested. The analysis of the electron density map reveals whether the chosen compound is bound to the protein. The aim of this chapter was to solve the first structure of WWP2 bound to an inhibitor to aid the design of more drug like WWP2 inhibitors. First, apo crystals of WWP2 that would diffract to a high resolution would need to be produced. Following from this, soaking conditions would be trialled in attempt to bind the ligand and retain crystal integrity. The quality of the data sets produced from these experiments would then be used to optimise soaking conditions with a chosen compound.

Initial efforts into the crystallisation of WWP2 was fraught with problems in the purification. Attempts to yield pure homogenous WWP2 HECT was unachievable due to the His tag cleavage problem. Zhang et al reported the purification of this construct, however the detail reported in the protein purification method was slightly brief. L. Wahl (2016) also reported

difficulties in the purification and crystallisation of WWP2 HECT. The crystallisation of the new WW2-2,3 linker HECT construct by Z. Chen (2017) sparked our curiosity as to whether this protein construct would be easier to purify. It was to our delight when sufficient quantities of pure protein could be produced and grown into diffractable crystals. The first round of crystal harvesting yielded apo crystals diffracting with a resolution of 2.74 Å and 2.55 Å, and soaked crystals with a resolution between 2.05- 3.58 Å. No extra density was picked up in the data sets from the soaked crystals. After increasing the DMSO concentration in an attempt to see the ligand, we obtained a data set of 1.99 Å resolution. Whilst there was no extra electron density that could be attributed to a ligand in the double difference map, we decided to use the data as a new improved search model. It was the generation of this higher quality search model which lead to additional electron density being picked out. Three data sets appeared to show the presence of NSC2805 bound (Figure 4.11). Two of the data sets have the ligand placed in one site, and one data set has the ligand placed in two different sites. All three structures have the ligand bound in the binding pocket which was predicted *in silico*. The additional site, located on the surface of the protein, came into vision when the concentration of the ligand was increased. The general quality of the protein ligand structures presented in this chapter leaves room for ambiguity. Where the ligand is placed in site 1, there is missing electron density around part of the phenyl ring. It is also concerning there is no density present around the two hydroxys in the upper ring. Where the ligand is placed in the second binding site, whilst the quality of the electron density fit is better (RSCC value of 0.75 compared to 0.61 in site 1), it is the nature of the site where the ligand is placed which brings about speculation. It is placed in the disordered part of the N-lobe, between a gap in the polypeptide chain which has not been modelled. Therefore there is a chance this electron density could just be a fragment of the polypeptide chain and not the ligand. There are limited meaningful interactions between the ligand and the protein at this site. There is only a 2.48 Å hydrogen bond between Gly486 which is a convincing strong interaction. The

other potential hydrogen bonds with Ser395 have a distance of 3.82 Å and 2.83 Å would exist as weak hydrogen bonds. Furthermore this site was not predicted *in silico*.

The first structure which has NSC2805 fitted in has the ligand placed almost exactly where it was predicted *in silico*. The presence of hydrogen bonds, π - π stacking and π -cation interactions around the ligand strengthen the argument that the electron density can be attributed to the ligand. Only experiments involving NSC2805 resulted in ligand bound structures. Protein ligand structures of pleurotin and purpurogallin, reported as inhibitors in chapter 1, were not obtained. Possible reasons for this could be the binding of these compound results in a conformational change of the protein, or that the protein requires a conformational change for binding to happen. If this were the case, co-crystallisation experiments need to be carried out. The purified protein is incubated with the compound for a set amount of time, at a given temperature, and then set up for crystallisation. As this may produce a different crystal form, it is more than likely alternate precipitant conditions would need to be screened. Alternatively, it could be a higher concentration of ligand is needed in order to see binding. Due to the compounds being dissolved in DMSO, there was a compromise in the amount of DMSO the crystal could tolerate and reaching high concentrations of compound. To alleviate this issue, alternate solvents could be used when dissolving the chosen compounds. Ciccone et al. (2015) reported a range of alternate solvents for soaking experiments. Crystals of WWP2 would be tested against these solvent mixtures to see what can be tolerated. If the crystals could tolerate higher concentrations of these alternate solvents than DMSO, then higher concentrations of the inhibitors could be reached. However this is assuming large amounts of solid can be obtained. In the case of some of the natural products tested in this chapter, this would be a significant challenge.

We felt there was a need to explore other techniques that would strengthen the structural information. Ligand NMR techniques are common place in drug discovery when developing

inhibitors. Therefore we shifted a focus onto using NMR to see whether we could validate a binding site and inform SAR.

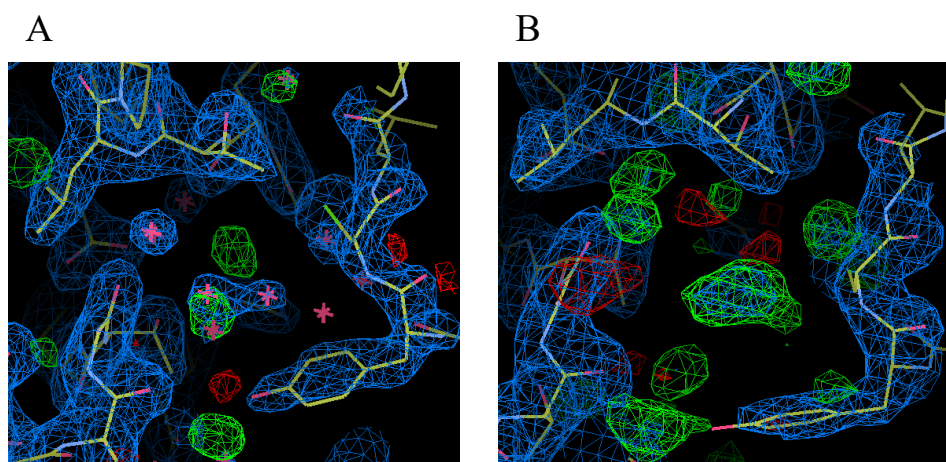


Fig. 4.11 -Electron density map comparison between apo (A) and soaked NSC2805 (B). The double difference is represented in blue and the single difference represented in green/red. The single difference density in B could not be attributed to water, and suggests weak density of the ligand is present.

Chapter 5

Utilising ligand NMR in the synthesis of isoalloxane based WWP2 inhibitors

5.1 Introduction

5.1.1 Discovery of novel WWP2 inhibitors by biochemical approaches

In 2018, we reported the discovery of the first WWP2 inhibitors (Watt et al., 2018). An ELISA based autoubiquitination assay was used to screen 1593 compounds from the NCI Diversity Set V and 450 compounds from the NCI Oncology Set V. The assay set up works by anchoring GST tagged WWP2 FL to glutathione coated wells. UbcH7 charged with flag ubiquitin is then added to the well and WWP2 undergoes autoubiquitination. The wells are then treated with an anti- flag HRP antibody conjugate. HRP has the ability to mediate the hydrogen peroxide dependent oxidation of TMB, resulting in a colour change which can be read at 450 nm. It is this read out which is used to measure autoubiquitination activity in the presence of potential inhibitors (Fig 5.1). The initial screen yielded 10 hits and subsequent concentration dependant assays calculated IC₅₀s ranging from 10.28-0.38 μ M.

The next step required confirmation of binding to the catalytic HECT domain via an orthogonal biophysical technique. STD and CPMG (Carr-Purcell-Meiboom-Gill) NMR experiments confirmed NSC2805, NSC650438, NSC369066, NSC3064, NSC194308 and NSC288387 to be binding to the HECT domain. The additional compounds which did not show binding by ligand NMR does not necessarily mean they are not binding to WWP2, as only the HECT domain is used for these experiments. Whilst it may be interesting to look at these compounds in the future, the main focus was to target the HECT domain and prevent its catalytic activity. The absence of unambiguous crystallographic data to elucidate binding mechanisms and inform analogue design of these compounds meant alternate techniques were required. The *in silico* docking of compounds represents the obvious alternative in the generation of structural data. The results from *in silico* docking can be strengthened by ligand NMR techniques.

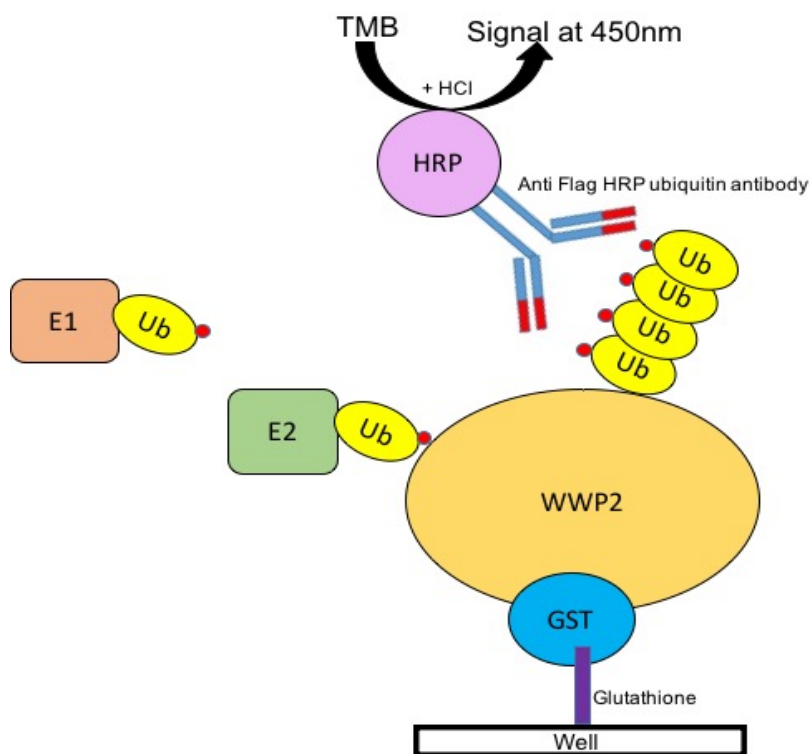


Fig. 5.1 -Schematic of the WWP2 autoubiquitination assay

5.1.2 Ligand NMR

5.1.2.1 STD Epitope maps

STD NMR (described in Section 3.1.1.3) can be used to produce ligand epitope maps (Mayer and Meyer, 2001), assisting in the validation of a computational derived protein ligand model. Ligand epitope maps reveal the binding mode of the bound ligand in the binding pocket. The ligand protons which are closer to the protein protons receive more saturation and the relative STD intensity can be analysed to build the epitope map. Epitopes are derived from the initial rate of STD build up (STD_0), which is calculated by fitting an exponential curve of saturation time against STD intensity. Once the curve has been fitted, the STD_0 for a selected proton can be calculated using the equation B shown in Figure 5.2. The STD_0 value is normalised

by dividing each value by the largest one. Mapping these values onto the ligand produces the ligand binding epitope map (Walpole et al., 2019).

$$(A) \quad STD(t_{sat}) = STD_{max} (1 - \exp(-k_{sat} t_{sat}))$$

$$(B) \quad STD_0 = \lim_{t_{sat} \rightarrow 0} \frac{\delta STD(t_{sat})}{\delta t_{sat}} = STD_{Max} k_{sat}$$

Fig. 5.2 -Equations used for calculating ligand epitope maps

5.1.2.2 DEEP STD NMR

Monaco et al. (2017) reported the development of a novel STD NMR technique named DiffErential EPitope mapping (DEEP), utilising the nature of surrounding amino acids to validate ligand binding poses, aiding further validation of an *in silico* model. Orientation of aliphatic, aromatic and polar amino acids around the binding sites can be determined. DEEP STD works on the basis of selectively saturating a particular type of proton environment and if that type of proton is present in the binding site, ligand protons close to those particular residues will receive more saturation relative to protons close to other residues. STD data sets are acquired at two different on-resonance frequencies. A comparison of aliphatic against aromatic residues is done by running an on-resonance experiment at 0.6 ppm (aliphatic) and a second on-resonance experiment at 6.55 ppm (aromatic). The DEEP STD factor for each proton is calculated by the STD intensity ratio of the two experiments, minus the average ratio. A positive factor indicates the ligand protons are close to an aliphatic region, whilst a negative factor indicates the protons are close to aromatics. Proximity to polar residues are calculated by use of differential solvents as opposed to differential saturation frequency. The first DEEP STD experiment is performed in D₂O and the second in H₂O. The ΔSTD is calculated as previously mentioned. Polar protons have the ability to exchange with the solvent, therefore the polar residues will contain deuterium in the D₂O experiment, which is

inactive in NMR and protons in the H₂O experiment, which are active in NMR. Interpretation of calculated Δ STD from these experiments means negative value indicates ligand protons are close to polar residues as the STD signal in the H₂O experiment would be stronger (Monaco et al., 2017)(Walpole et al., 2019).

$$\Delta STD_i = \frac{STD_{i,1}}{STD_{i,2}} - \frac{1}{n} \sum_i^n \left(\frac{STD_{i,1}}{STD_{i,2}} \right)$$

Fig. 5.3 -Equation used to calculate the DEEP STD factor

5.1.3 Experimental aims

Early hit- to- lead development of compounds in drug discovery is when initial hits are improved to have higher affinity and greater selectivity to the target. By using the techniques described, a computational derived NMR validated protein ligand model can be established. NSC288387 was picked out from the 10 initial WWP2 inhibitor hits for further analysis as it gave the clearest STD signals, suggesting a clear binding epitope map could be built and used to inform analogue development. NSC288387 is a flavin compound consisting of a tricyclic isoalloxazine core with a *N*-substituted phenyl ring.

The NMR validated *in silico* model can then to be used to drive analogue design and synthesis, with the aim of bringing the ligand activity down to the sub micromolar range. IC50s of the synthesised analogues are to be initially generated for WWP2. This information can be used to investigate SAR for a putative binding site. Activity is then to be tested against Nedd4 to test for specificity.

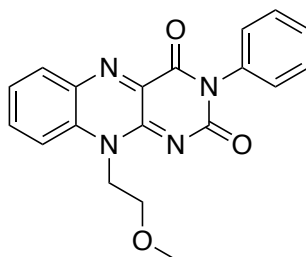
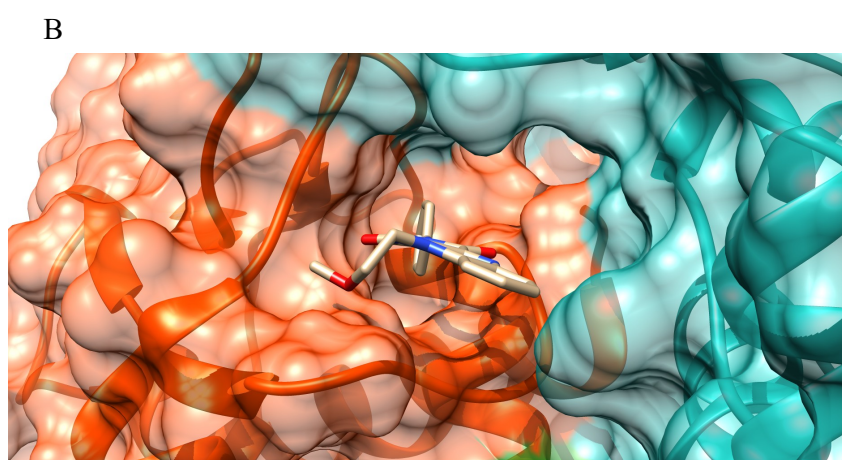
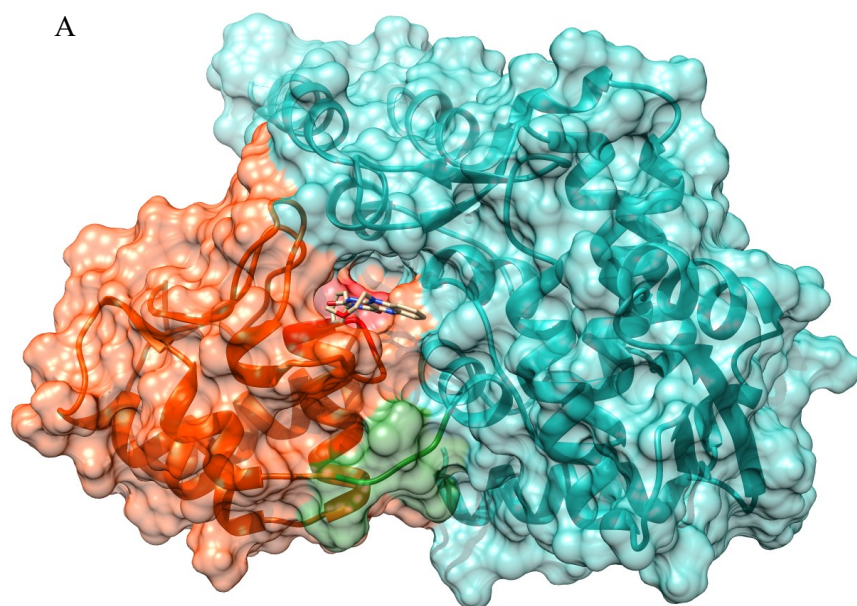


Fig. 5.4 -Structure of NSC288387

5.2 Results

5.2.1 *In silico* modelling of NSC288387 into WWP2 HECT

The WWP2 HECT (aa 486-865) crystal structure (PDB:4Y07) was used as the model for docking NSC288387. NSC288387 was docked into the 1st site as described in Section 4.2.1. We predicted this to be the most likely binding region due to the mechanistic relevance the site holds. As the site lies close to the hinge loop sitting in the interface of the N and C lobe, it was theorised targeting this region would hinder the HECTs intrinsic flexibility required for transferring ubiquitin from the N-lobe bound E2 to the catalytic cysteine in the C-lobe. The lowest energy solution is shown in Figure 5.5. All docking calculations were performed in collaboration with Dr Jesus Angulo, Dr Sam Walpole and Dr Serena Monaco in the School of Pharmacy at UEA. The docking solution shows the compound sitting in the interface of the N and C lobe, close to the hinge loop. The surrounding residues are His569, Tyr587, Asp623, Leu795, Thr799, Thr801, Arg803, Pro805, Val806, Glu811, Leu812, Ile813 and Asn840. There is a 1.9 Å hydrogen bond from Val806 to the methoxy on NSC288387. The carbonyl on the N substituted ring makes a 2.28 Å hydrogen bond to Arg803. There is potential for π - π stacking from Tyr587 and the N substituted phenyl ring, however at a distance of 4.61 Å, this is stretching beyond the limit of conventional π - π interaction distance (Brylinski, 2018).



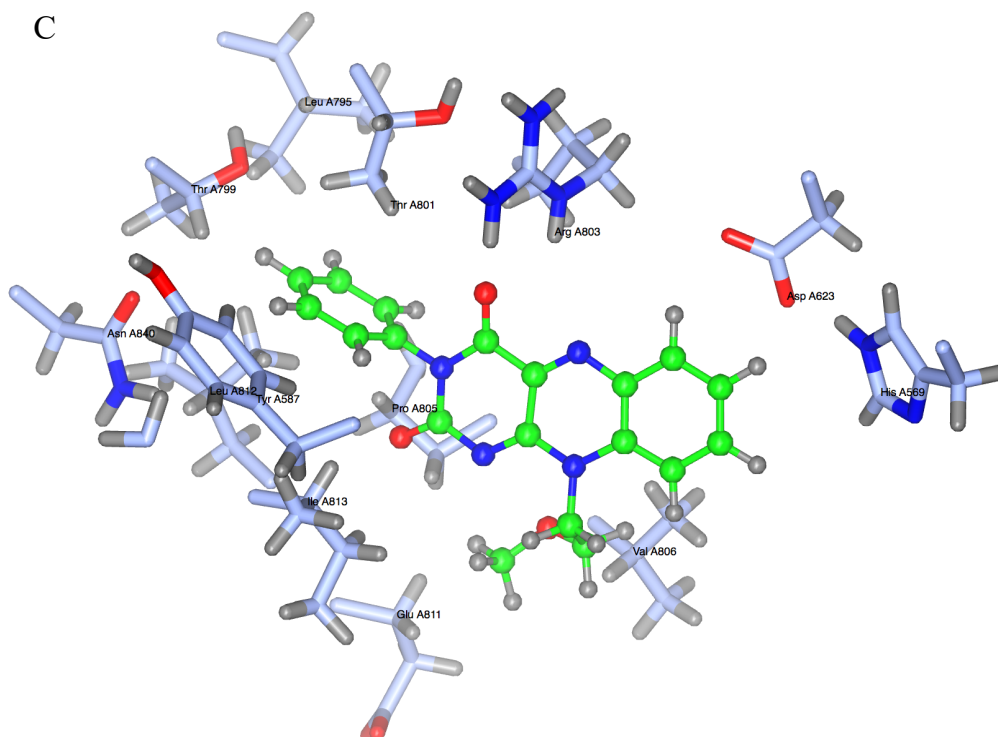


Fig. 5.5 -NSC288387 docked into WWP2 HECT. The N and C lobe are shown in blue and orange respectively, the hinge loop shown in green. (B) Closer view of NSC288387 sitting in the predicted binding site, the *N*-substituted phenyl ring is facing in towards the central cavity of the protein. (C) Labelled residues located around the compound.

5.2.2 NMR validation of the WWP2:NSC288387 *in silico* model

Ligand NMR was then used to assess the viability of this proposed model of binding. WWP2 HECT was expressed and purified as stated in Section 3.2.1.1. The purified protein was exchanged into 25 mM deuterated Tris-Cl pH 8.9, 100 mM NaCl and 1 mM deuterated DTT made up in D₂O. All ligand NMR experiments were run with 20 μ M protein and 100 μ M ligand. Binding epitope mapping and DEEP STD NMR analysis was performed in collaboration with Dr Jesus Angulo, Dr Sam Walpole and Dr Serena Monaco. The binding epitope map was obtained by normalising the STD signals (%) against the strongest signal (Ha). The comparison of the binding epitope map with the structural model is in excellent agreement. The *N*-substituted phenyl ring, which is predicted to be sitting deeper into the

pocket, shows to have the strongest interactions. The aliphatic ether chain shows weaker interactions, which would suggest the chain is solvent exposed, this is clearly evident in the structural model.

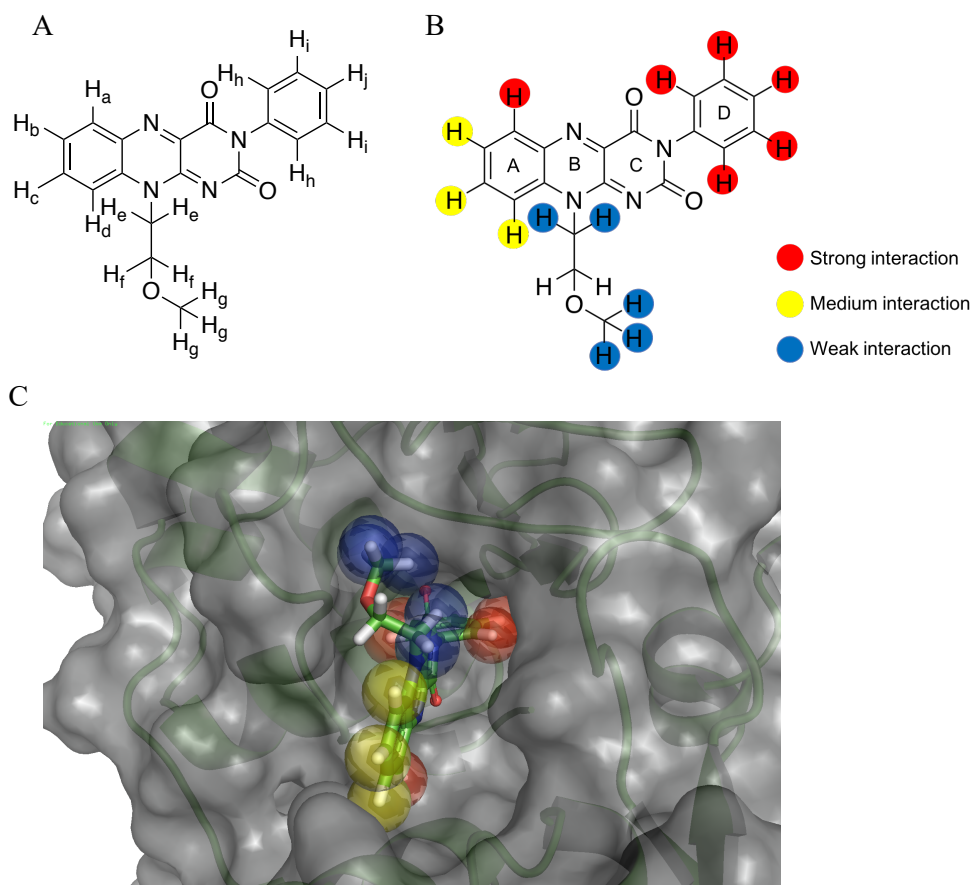


Fig. 5.6 -Ligand epitope map of NSC288387 bound to WWP2 HECT (A) Labelled hydrogens of NSC288387. (B) 2D schematic of the ligand epitope map (Adapted from Watt et al.). (C) 3D representation of the ligand epitope map in context with the predicted binding site.

DEEP STD data was obtained by irradiating WWP2 HECT at 0.5 ppm and 1.5 ppm. Irradiation of the protein at frequency at the 0.5 ppm corresponds to the methyl groups of valine and isoleucine, and irradiation at the 1.5 ppm frequency corresponds to the β/γ arginine protons. These irradiation frequencies were chosen due to the presence of Arg803 and Val806 in close proximity to ring A. Both these residues were singled out as they could be targeted without overlapping with the ligand frequencies, such as aromatic frequencies. The

Hydrogen	STD%	Normalised STD %
a	15.51	100
b	11.57	75
c	11.02	71
d	10.09	65
e	4.95	32
f	-	-
g	7.53	49
h	13.79	89
i	12.63	81
j	13.01	84

Table 5.1 -WWP2 HECT: NSC288387 STD epitope values. Relative intensity descriptions are given as 80-100 %= strong, 60-79 %= medium, <60 %= weak.

DEEP STD factor for each hydrogen was calculated as stated in section Section 5.1.2.2 where 0.5 ppm and 1.5 ppm is Exp1 and Exp2, respectively. Positive DEEP STD factors correspond to interactions at the 1.5 ppm frequency and negative values correspond to interactions at 0.5 ppm frequency. Protons a and b gave a positive signal, and inspection of the model reveals those protons to be orientated towards Arg803. Protons c, d and e are predicted to be orientated towards Val803, and they satisfyingly all give negative DEEP STD factors.

Hydrogen	STD % 0.5 ppm	STD % 1.5 ppm	DEEP-STD factor
a	13.66	15.66	0.258
b	11.35	12.76	0.236
c	9.25	5.46	-0.298
d	9.9	6.2	-0.262
e	6.08	4.28	-0.183
f	-	-	-
g	7.61	6.99	0.030
h	12.39	11.46	0.037
i	13.53	14.2	0.161
j	13.14	11.92	0.019

Table 5.2 -WWP2 DEEP STD factors

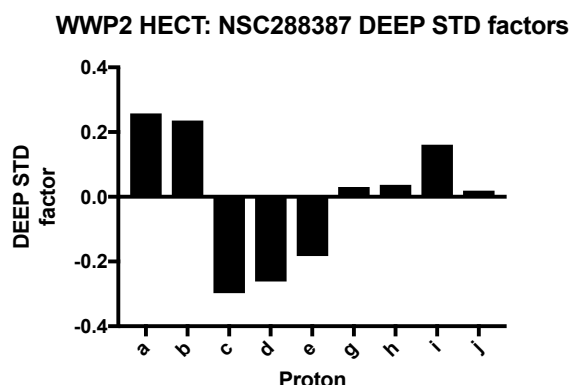


Fig. 5.7 -DEEP STD factors for 288387 bound to WWP2 HECT

5.2.3 *In silico* modelling of NSC288387 into WWP1 HECT

WWP1 HECT shares 83.6 % (Uniprot alignment) sequence homology with WWP2 HECT. For this reason, it is unsurprising that the putative binding pocket is also present in WWP1. To investigate whether NSC288387 is predicted to bind differently in WWP1, it was docked into the WWP1 HECT domain (pdb: 1ND7). The WWP1 binding cleft is narrower than in WWP2, requiring an induced fit docking method. NSC288387 has the same bound orientation as predicted in WWP2, with the N-phenyl ring pointing towards the central cavity. The surrounding residues are Asn638, Tyr639, Cys640, Asp675, Thr851, Thr853, Arg855, Met865 and Asn892. There are hydrogen bonds between Thr853, Cys640 and Asn638 with the carbonyls on the tricyclic core. Arg855 makes a hydrogen bond with the oxygen in the methoxyethyl side chain. Tyr639 forms a π - π stacking interaction with ring A of the tricyclic core. Arg855 has the potential to form a π -cation stacking interaction with the aromatic rings.

5.2.4 NMR validation of the WWP1:NSC288387 *in silico* model

As with WWP2, ligand NMR was used to validate the proposed model. The DNA sequence corresponding to the WWP1 HECT domain (aa 546-917) lacking the 5 C-terminal residues

was cloned into the pET28a vector. The protein was expressed as an N-terminal hexa-His tag construct. The protein was purified using Ni-NTA column and was dialysed into 5mM Tris Cl pH 8.0, 5 mM DTT. The protein was exchanged into the deuterated version of buffer for NMR experiments. As with WWP2, 20 μ M of HECT to 100 μ M ligand was used. The binding epitope shows the *N*-phenyl ring to have the strongest interactions, consistent with the prediction of the ring sitting deeper in the pocket. The more solvent exposed part of the tricyclic core and the methoxyethyl chain have weaker interactions.

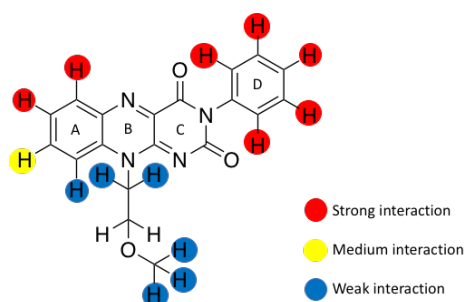


Fig. 5.8 -Ligand epitope of NSC288387 when bound to WWP1 HECT

Hydrogen	STD%	Normalised STD %
a	45.37	100
b	40.35	89
c	31.01	68
d	26.68	59
e	23.59	52
f	-	-
g	28.12	62
h	42.79	94
i	44.49	98
j	44.06	97

Table 5.3 -WWP1 HECT: NSC288387 STD epitope values. Relative intensity descriptions are given as 80-100 %= strong, 60-79 %= medium, <60 %= weak.

DEEP STD data was obtained by irradiation of WWP1 HECT at 0.5 ppm and 1.5 ppm. The DEEP STD factor for each hydrogen was calculated in the same manner as WWP2 in Section 5.2.2. The DEEP STD factors showed good agreement with the docked model. Proton b and c on ring A have positive values and are close to Arg855. Protons d, e and g have negative values and the model shows the aliphatic chain pointing towards Leu855. Even though the DEEP STD factors are in good agreement, their values are much lower than compared to WWP2. We could not validate the docked model with as much confidence as WWP2 though inspection of the model predicts these irradiated residues to be further from the ligand than in WWP2, giving a possible reason as to why the values are particularly low.

Hydrogen	STD % 0.5 ppm	STD % 1.5 ppm	DEEP-STD factor
a	48.57	47.63	-0.0163196
b	41.96	41.96	0.05862373
c	34.58	36.25	0.10691755
d	34.18	28.12	-0.1186729
e	27.58	23.14	-0.1023625
f	-	-	-
g	33.2	29.18	-0.0624606
h	48.57	44.93	0.03927022
i	50.51	49.05	0.02971857
j	47.63	47.17	0.04896595

Table 5.4 -WWP1 DEEP STD factors

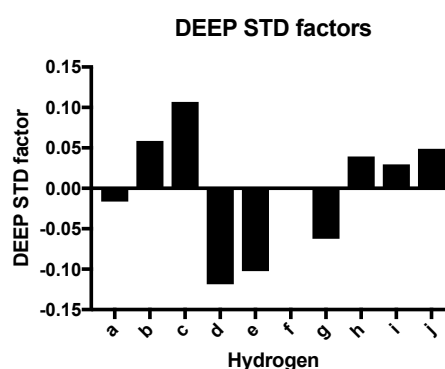


Fig. 5.9 DEEP STD factors for 288387 bound to WWP1 HECT

5.2.5 *In silico* modelling of NSC288387 into Nedd4 HECT

NSC288387 was docked into the putative binding site between the N and C lobe of Nedd4 HECT. As with WWP1 and WWP2, the binding pose is analogous with the *N*-phenyl ring pointing inwards. The surrounding residues are Phe594, Asp653, Phe655, Leu780, Glu818, Arg822 and Gln825. Arg822 and Gln825 form hydrogen bonds with the carbonyls on ring C. Phe594 is orientated to make π - π stacking interactions and Arg822 is in a position to form π -cation interactions. As with the WWP1 and WWP2 HECT domains, we wanted to validate this model by STD epitope maps and DEEP STD. The DNA sequence corresponding to the Nedd4 HECT domain was cloned in the pET28a plasmid. Trial induction conditions were screened across Rosetta (DE3), BL21 (DE3), BL21 Star (DE3), CodonPlus RP and pLys at 20 °C, 25 °C and 30 °C. The optimal condition which gave the highest Nedd4 HECT expression relative to unwanted bacterial proteins was 30 °C overnight in the BL21 Star (DE3) cell line. The protein was expressed with a *N*-terminal 6His- tag and was purified as described in Section 2.1.2.3. Nedd4 HECT was dialysed into HEPES buffer. Due to difficulties in concentrating, we were unable to buffer exchange into deuterated buffer. Therefore, STD experiments were run in 50:50 HEPES: Tris in D₂O at a protein concentration of 20 μ M. We were unable to generate interpretable data from these conditions.

5.2.6 Design of NSC288387 analogues

Owing to the fact that the ligand NMR data suggested the *N*-aryl ring had the strongest contacts and was sitting deepest into the putative binding site, it seemed the most logical place to make alterations that would have an effect on activity. A range of substitutions on the ring were designed and docked into the binding site (Table 5.5). The selected compounds were ranking according to the addition of any positive interactions. The extension of the *N*-phenyl ring into the pocket by the insertion of a short alkyl chain had the best predicted

binding. *Para* substitution ranks higher than *ortho* and *meta* substitution, with the *para* substitutions causing the compound to sit in a different orientation, deeper into the pocket. A synthetic methodology was devised which would make the isoalloxazine core, from which various *N*-substitutions could be made. The first strategy to synthesise the isoalloxazine was devised according to Carell at al. Starting from 2-nitroaniline, the amino would be protected with a trifluoro acetyl group using trifluoroacetic anhydride. The amine would then be alkylated via an S_N2 reaction and the protecting group removed in a single step. The reduction of the nitro and ring cyclisation using alloxane monohydrate would happen in a one pot two step reaction. This would give rise to the tricyclic isoalloxazine core, leaving the ability for N-C coupling reactions to afford the N-(hetero)aryl analogues.

Rank	Substituent	Rank	Substituent
1		8	
2		9	
3		10	
4		11	
5			
6			
7			

Table 5.5 -Ranked analogues

5.2.7 Organic synthesis

5.2.7.1 Amine protection and alkylation

Compound **12** was first protected with a trifluoroacetyl group. The aniline undergoes a nucleophilic addition-elimination reaction with trifluoroacetic anhydride. The reaction was stirred in dichloromethane at 0 °C for 2 hours to afford **13** in 87 % yield. The alkylation of the NH was achieved by heating **13** with excess 1-bromo-2-methoxyethane, in the presence of potassium carbonate at 100 °C overnight, to give the desired product in 56 % yield. Mixtures of mono alkylated and di alkylated products were obtained and separated. We tried directly

alkylating **12** to **14** in the same conditions. However, LCMS only showed a 1 % conversion to the product. Due to the start of the synthesis already being relatively inefficient, an alternative method to make compound **14** was carried out.

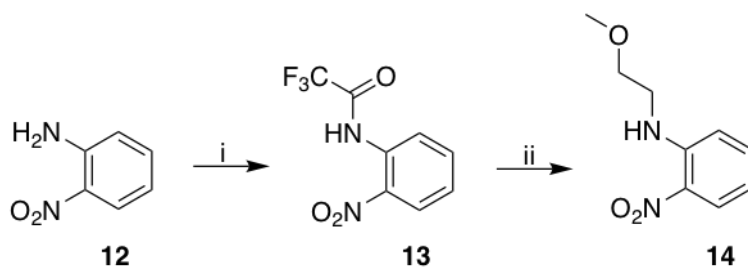


Fig. 5.10 -Scheme 7. Reaction conditions: (i) $(\text{CF}_3\text{CO})_2\text{O}$, DCM, 0°C for 2 hours, 87 %. (ii) 1-bromo-2-methoxyethane, K_2CO_3 , DMF, 100°C overnight, 56 %.

5.2.7.2 Nucleophilic aromatic substitution

Aryl rings with electron withdrawing groups ortho or para to a leaving group can undergo nucleophilic aromatic substitution via the $\text{S}_{\text{N}}\text{Ar}$ mechanism. Ortho and/or para electron withdrawing groups, such as the nitro group, are necessary in the formation of the resonance stabilised Meisenheimer complex prior to the elimination of the leaving group. As the methoxyethylamine group in **14** is ortho to the nitro, we could start from fluoronitrobenzene and perform an $\text{S}_{\text{N}}\text{Ar}$ reaction with 2-methoxyethylamine to produce **14** in 90 % yield.

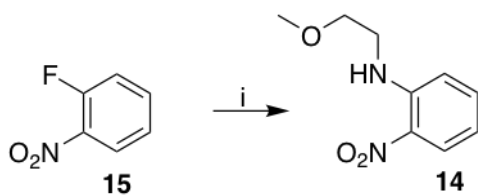


Fig. 5.11 -Scheme 8 Reaction conditions: (i) 2-methoxyethylamine, NaCO_3 in ethanol, 80°C overnight, 90 %.

5.2.7.3 Reduction of nitro group and condensation with alloxan monohydrate

The reduction of the **14** to **16** was done by hydrogenation with Pd/C catalyst. Once the complete conversion to the aniline was observed by LCMS, the condensation reaction with alloxan monohydrate was carried out without isolation of the aniline. The tricyclic isoalloxine core **17** was isolated in 53 % yield. The reaction was assumed to be regioselective to give the single isomer **17** by literature precedent (Epple et al., 1997) (Lacroix et al., 1987).

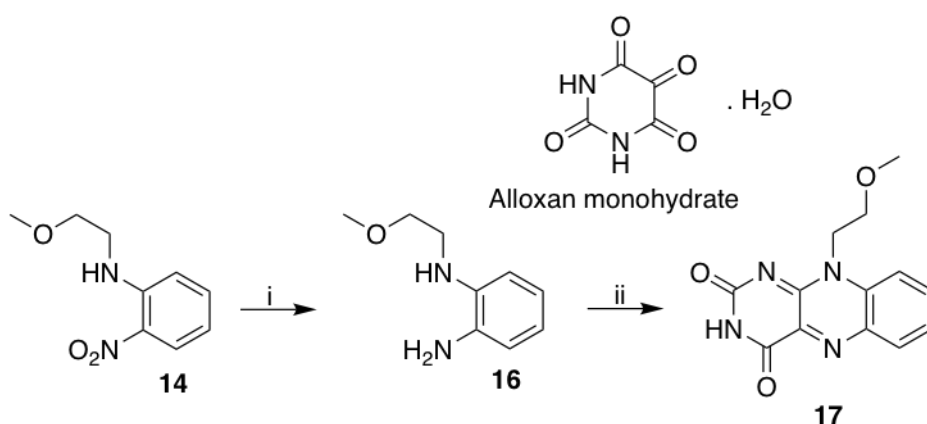


Fig. 5.12 -Scheme 9. Reaction conditions: (i) Reduction of nitro to amine. H₂, Pd/C in acetic acid, material was carried through to the next synthetic step without further purification. (ii) Ring formation. Boric acid and alloxan monohydrate in acetic acid, 53 % yield.

5.2.7.4 Chan- Lam coupling

Attenberger et al.(2008) reported the *N*-arylation of flavins using a method under Chan- Lam conditions. First reported in the simultaneous publications by Chan and Lam, the Chan- Lam reaction is a cross coupling reaction between aryl boronic acids and amines, catalysed by copper acetate. Unlike conventional transition metal cross coupling which consists of a nucleophile and an electrophile (such as organoboron and an organohalide), the Chan- Lam couples two nucleophiles, organoboron and an amine (Fig 5.13). The Chan- Lam coupling is an appealing route to N-C coupling owing to its mild conditions. Initially, the phenyl boronic acid was coupled with **17** in 57 % yield. Giving the moderate yield, we proceeded with the coupling of range of substrates which represented the scope of the designed compounds in Table 5.5. A summary of substrates tested with yields is shown in Table 5.6. Yields varied from 4-57%, with the majority of yields <20% and a number of substrates not working. The highest yield came from the unsubstituted phenyl boronic acid, any additional substitutions dramatically lowered the yield. Despite the low yields, at least 20 mg of product from each successful reaction was obtained, when only approximately only 1 mg was need for biological testing. Reactions with substituents in the *ortho* position did not work.

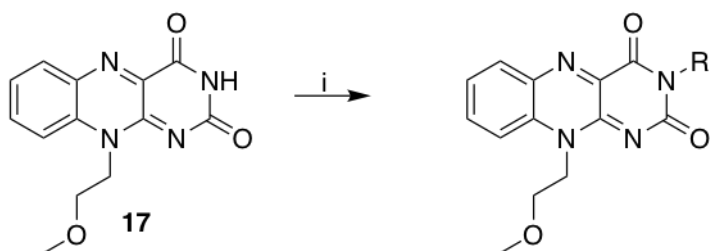


Fig. 5.13 -Scheme 10. Reaction condition: (i) R-B(OH)₂, Cu(OAc)₂ and triethylamine in DMF.

Compound	R	Yield %
18		57
19		19
20		6
21		4
22		5
23		17
24		13
25		11

Table 5.6 -Chan Lam substrate yields

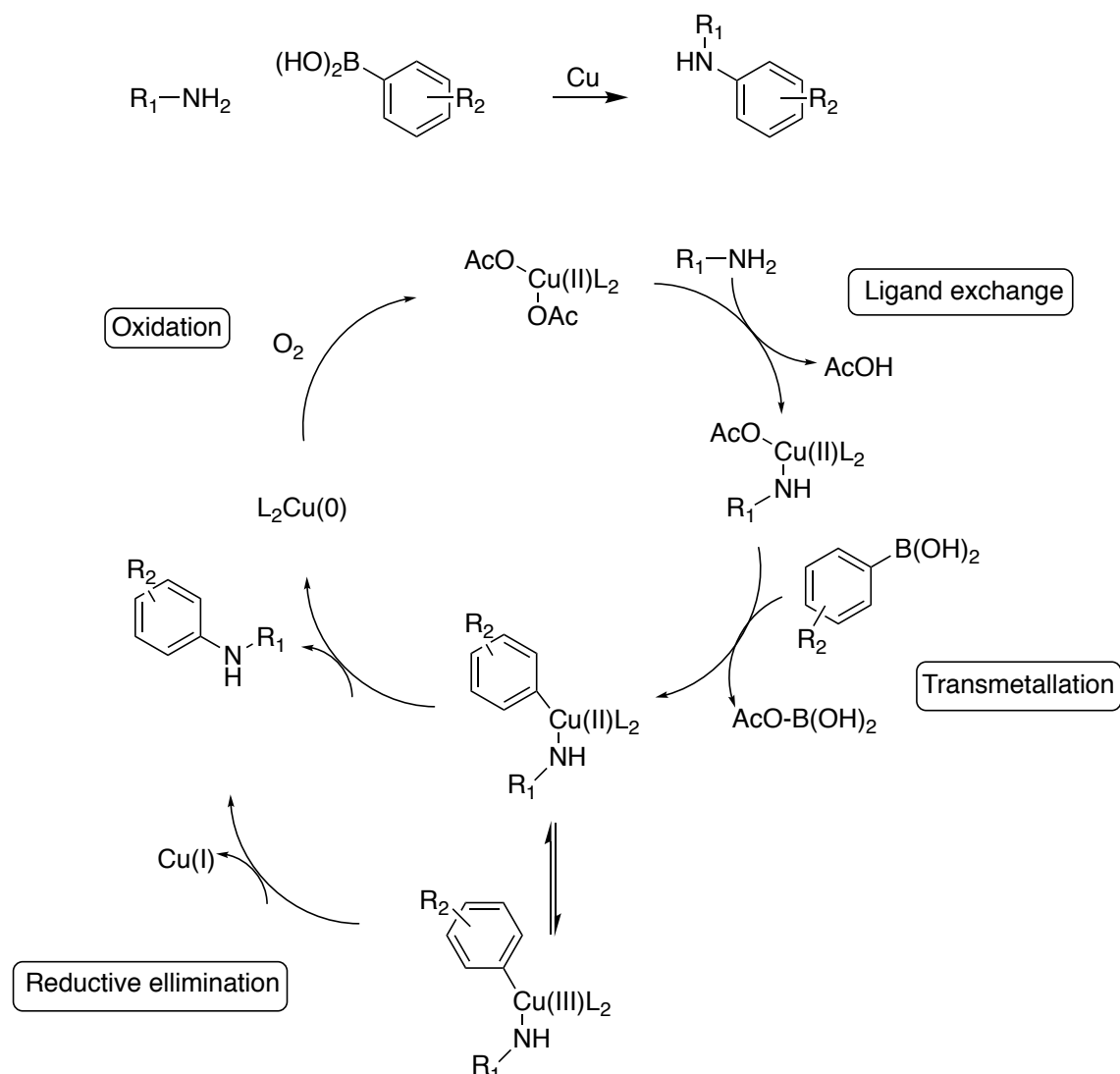


Fig. 5.14 A simplified Chan Lam catalytic cycle. The copper species complexes with the amine and boronic acid through ligand exchange and transmetalation. There was significant ambiguity as to whether the coupling by reductive elimination proceeds through Cu(II) Cu(0) or Cu(II) Cu(III) Cu(I) Cu(0). Recent mechanistic studies suggests it goes through the latter (Vantourout et al., 2017).

The NSC288387 WWP2 *in silico* model shows a large amount of space surrounding ring A. Both Arg803 and His569 can make aromatic interactions with NSC288397, therefore substituents on ring A could have an impact on the strength of this interaction. We therefore attempted to design in electron activating and deactivating groups. Hydroxyl and trifluoromethyl were chosen as they would not only add additional hydrogen bonds, but also

alter the π -cation interaction. 3-Fluoro-4-nitrophenol underwent an S_NAr reaction with 2-methoxyethylamine to give compound **27** in 61 % yield.

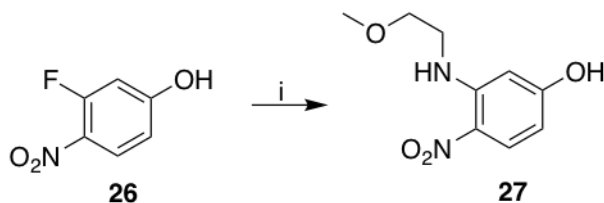


Fig. 5.15 -Scheme 11. Reaction conditions: (i) 2-methoxyethylamine, Na_2CO_3 in ethanol, 80°C overnight, 61 %.

However, attempting to reduce the nitro and form the tricyclic core with the conditions used previously, did not work.

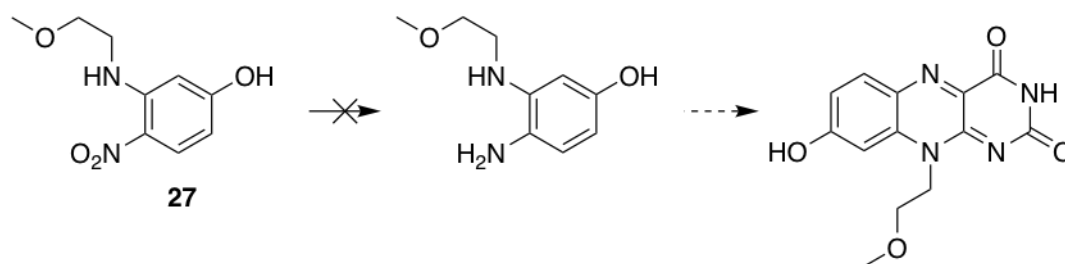


Fig. 5.16 -Scheme 12. Failed reduction and cyclisation.

We did however have much better success with the trifluoromethyl analogue. 4-Fluoro-3-nitrobenzotrifluoride was reacted with 2-methoxyethylamine to give compound **29** in 86 % yield.

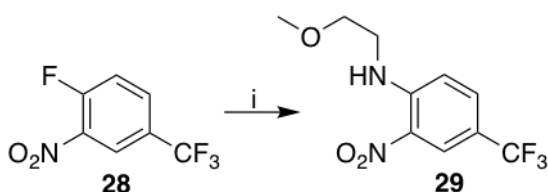


Fig. 5.17 -Scheme 13. Reaction conditions: (i) 2-methoxyethylamine, NaCO_3 in ethanol, 80°C overnight, 86 %.

The reduction of the nitro by H_2 over Pd/C was monitored by LCMS until completion, but as with **16**, the intermediate amine was not isolated and the crude was reacted with alloxane monohydrate to give **31** in 9 % yield.

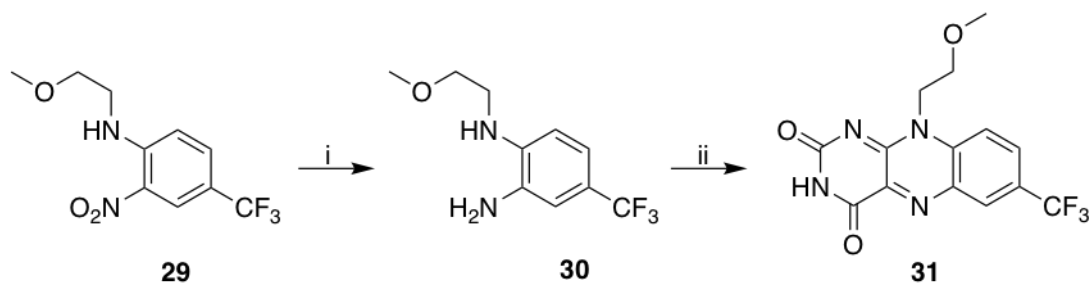


Fig. 5.18 -Scheme 14. Reaction conditions: (i) H_2 , Pd/C in acetic acid, material was carried through to the next step without further purification. (ii) Boric acid and alloxan monohydrate in acetic acid, 9 % yield.

The Chan Lam coupling of compound **31** with phenylboronic acid gave the target compound **32** in 64 % yield.

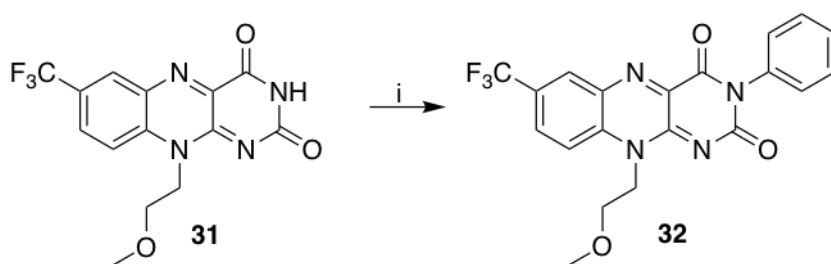


Fig. 5.19 -Scheme 15. Reaction conditions: (i) Benzene boronic acid, $\text{Cu}(\text{OAc})_2$ and triethylamine in DMF, 64 % yield.

5.2.8 Testing of analogues against WWP2 autoubiquitination

NSC288387 analogues were tested for *in vitro* WWP2 autoubiquitination inhibitory activity by using the assay reported in Watt et al. Concentration dependent assays with compounds at 100 μM , 50 μM , 25 μM , 10 μM , 5 μM and 1 μM at 1 % DMSO-d6 were performed. Levels of autoubiquitination were measured as the relative activity percentage of the full autoubiquitination reaction. Curves were generated using the non linear regression of inhibitor concentration versus normalised response. The addition of a chloro substituent meta to the isoalloxazine ring produced the greatest improvement in activity. NSC288387 had an IC_{50} of 2.3 μM , whilst the addition of the chloro increased the activity 10 fold to 0.23 μM . Removal of the phenyl ring results in loss of activity with both the original tricyclic core and the trifluoromethyl substituted version. Removal of the ring in both versions decreases the potency by at least 10 fold. Adding the strongly deactivating trifluoromethyl group to the A ring of the isoalloxazine core decreased its activity approximately 3 fold to 6.68 μM . The electron donating groups methyl and methoxy had negative effects on the original compounds activity, whilst the electron withdrawing groups chloro, fluoro, nitrile, methyl ester improved activity. It was assumed addition of electron withdrawing groups in the *para* position were having a positive effect. However, the most strongly electron withdrawing group, the trifluoromethyl, reduces the compounds potency down to 7.43 μM .

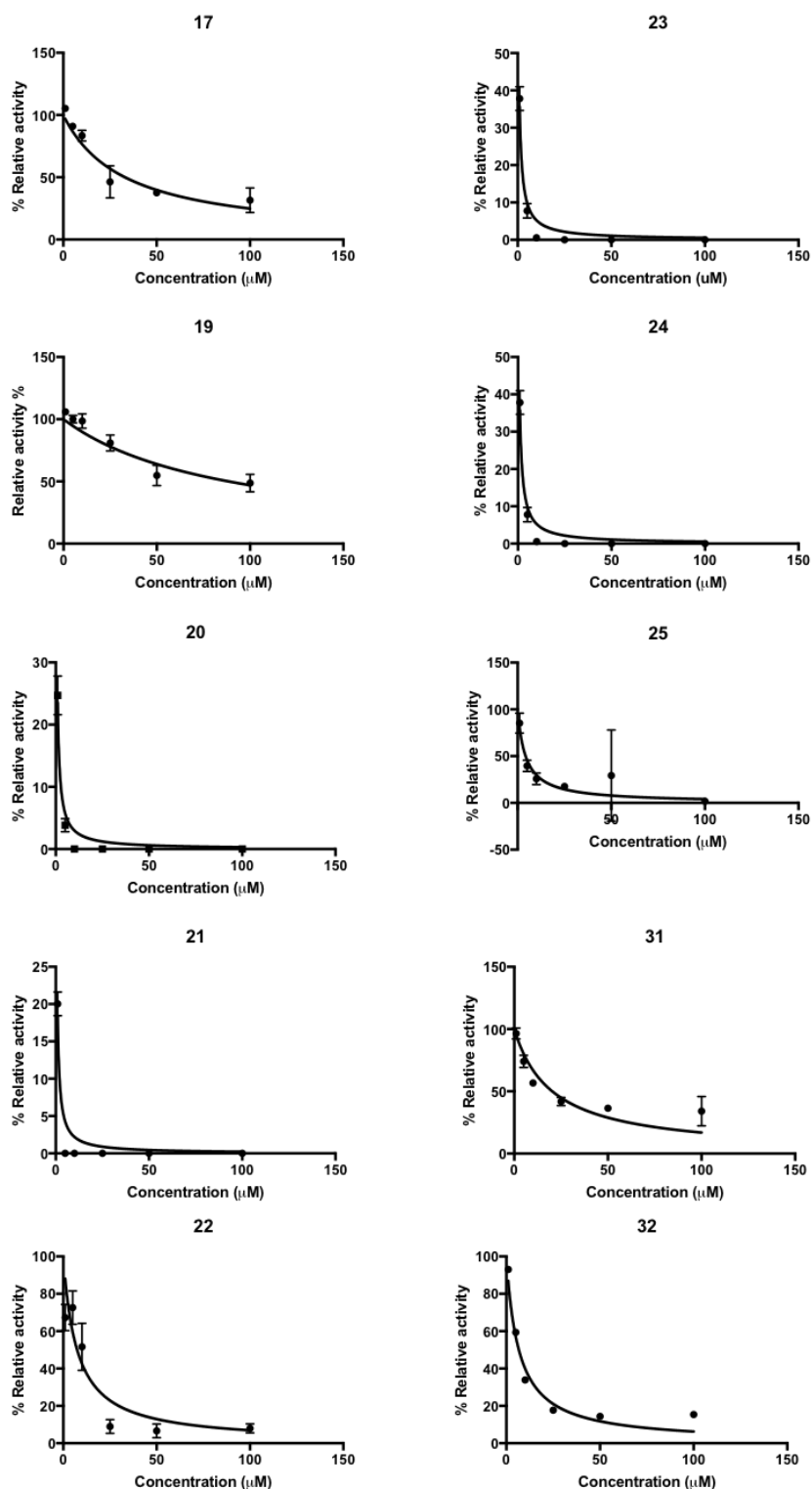
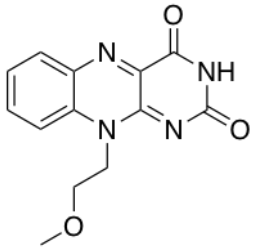
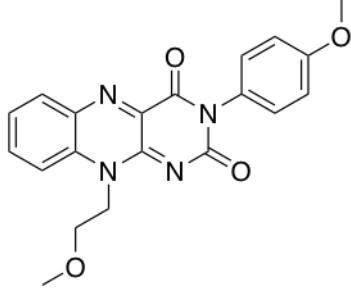
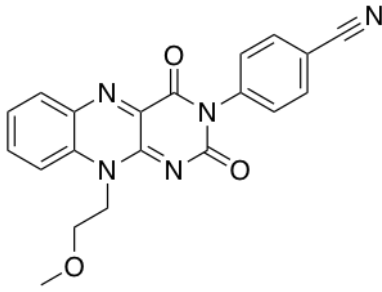
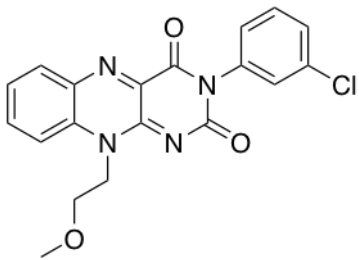
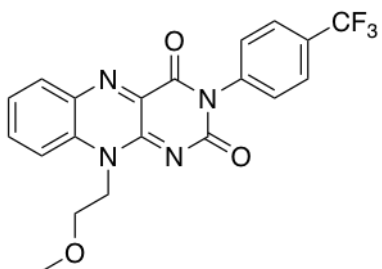


Fig. 5.20 -WWP2 autoubiquitination IC₅₀ curves

Name	Structure	IC50 μ M
17		33.06
19		117.8
20		0.31
21		0.23
22		7.43

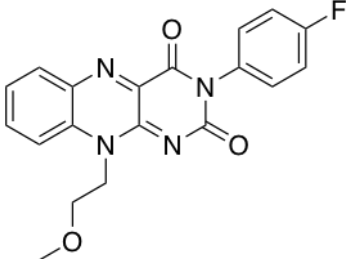
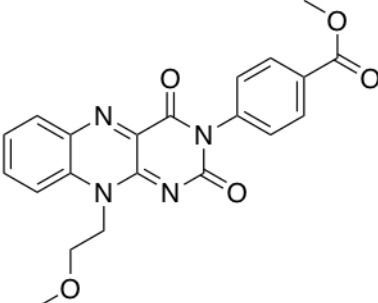
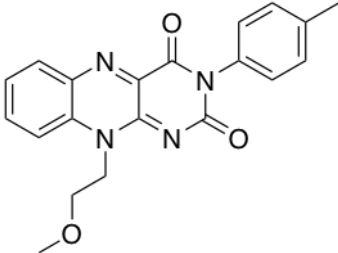
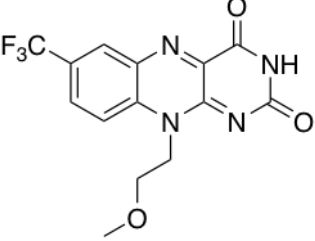
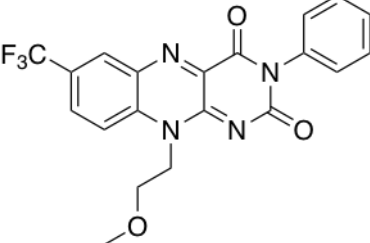
Name	Structure	IC50 μ M
23		0.55
24		0.55
25		4.26
31		20.31
32		6.68

Table 5.7 -NSC288387 analogue WWP2 autoubiquitination IC50's

5.2.9 Testing of compounds against Nedd4

To determine specificity and to help identify lead analogues, it was important to test the compounds against other Nedd4 ligases. Assays testing in vitro autoubiquitination for WWP1 and Nedd4 has previously been developed in the Chantry lab. The Nedd4 autoubiquitination assay is analogous to the WWP2 assay, with GST-Nedd4-FL being used. IC₅₀s were generated in the same way as for WWP2 stated in Section 5.2.8. The most potent analogue is the methyl ester variant with an IC₅₀ of 0.6 μ M. The least potent analogue is the methoxy variant with an IC₅₀ of 88.57 μ M. The analogues follow a similar trend as was seen with WWP2, where generally deactivating substituents improve the potency and activating substituents decrease the potency. As with WWP2, removal of the N-phenyl ring has a negative effect on activity. Addition of the trifluoromethyl group on ring A of the tricyclic core has little effect on the IC₅₀. The improvement in IC₅₀s is lower compared to WWP2.

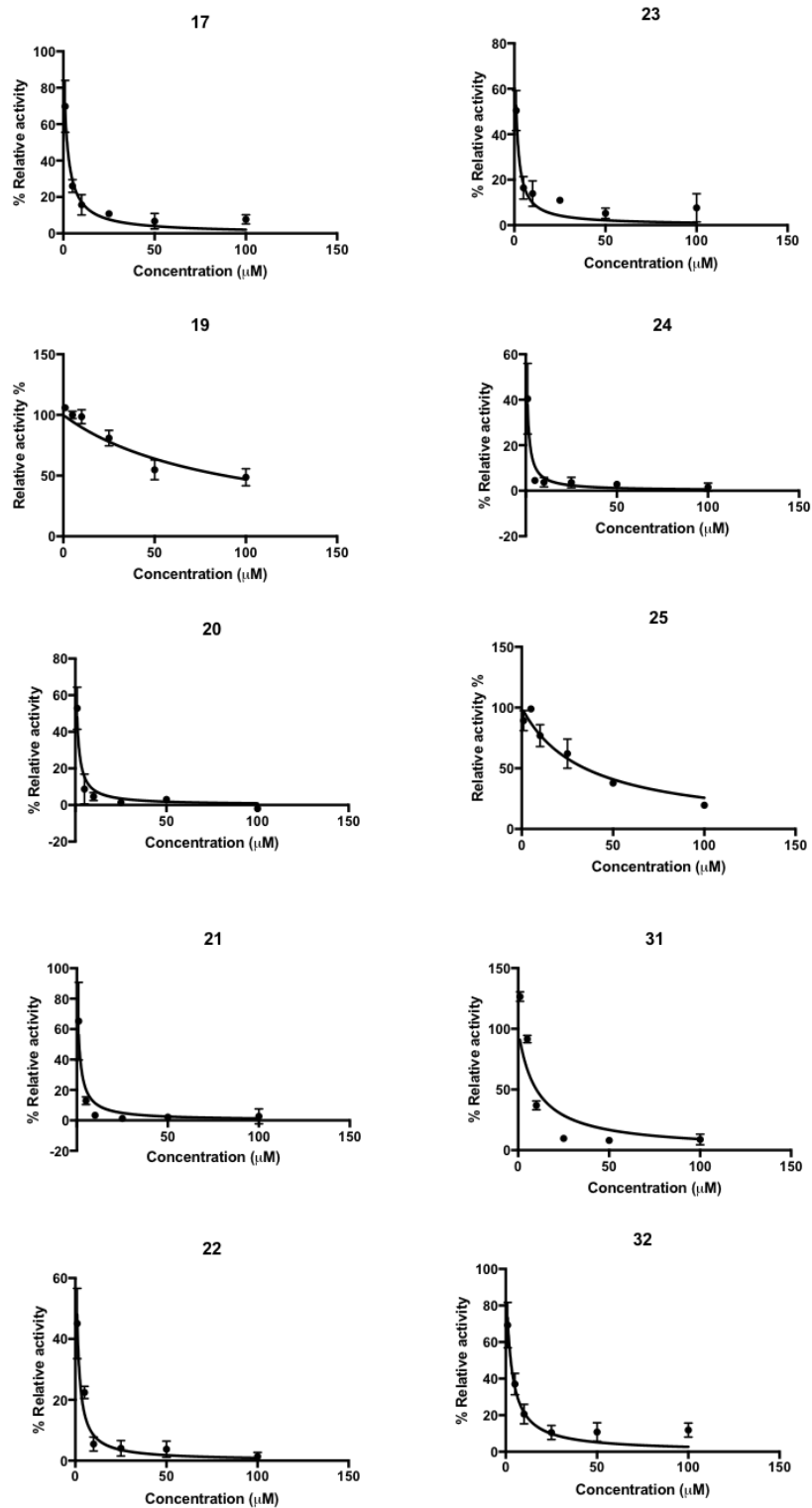
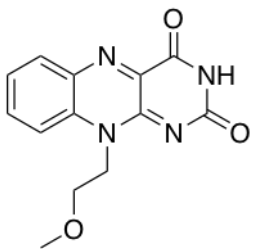
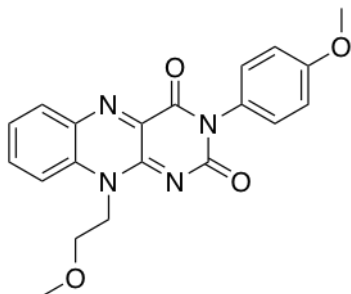
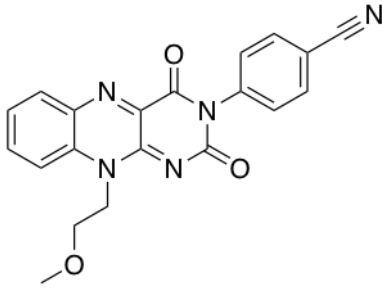
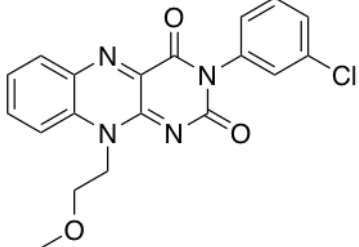
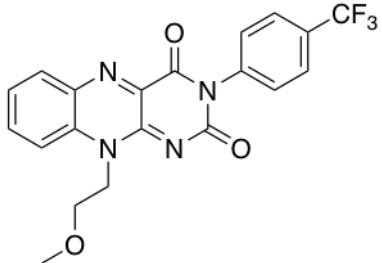


Fig. 5.21 -Nedd4 autoubiquitination IC50 curves

Name	Structure	IC50 μ M
17		2.08
19		88.57
20		0.92
21		1.29
22		0.92

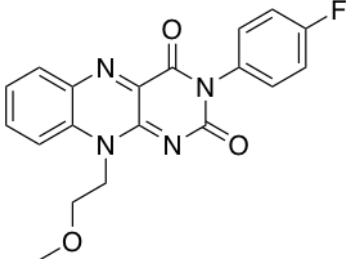
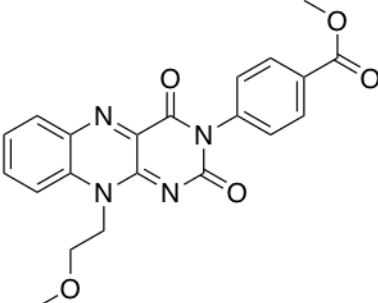
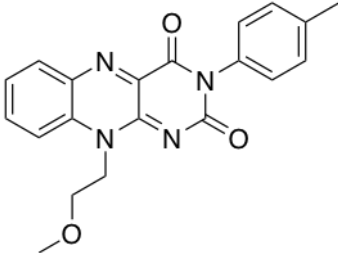
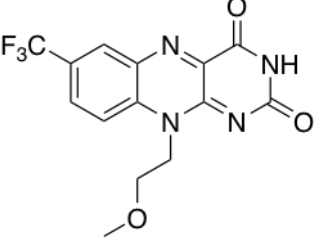
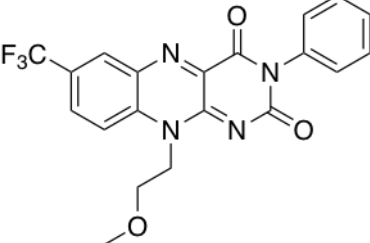
Name	Structure	IC50 μ M
23		1.1
24		0.60
25		34.6
31		10.05
32		2.74

Table 5.8 -NSC288387 analogue Nedd4 autoubiquitination IC50's

5.3 Discussion

Gaining insight as to where inhibitors bind to proteins is a crucial part of the drug design process. To improve aspects such as potency and specificity, one must know the nature of the environment around the inhibitor. Once this is known, modifications can be designed into the selected compound. Neighbouring amino acids can be targeted for selected non covalent interactions in order to improve potency. The physical space around the compound can be viewed and used to fill with desirable moieties which would improve a compounds characteristics. Crystallographic data of a compound bound to its target is the pre-eminent method to achieve this. However in the absence of unambiguous crystal data, other methods must be used in order to distinguish how an inhibitor is binding. The aim of this chapter was to elucidate a binding site in the WWP2 HECT domain which can be targeted by small molecules, with limited use of crystallographic data. Once this had been achieved, modifications to NSC288387 were made with the intention of improving potency and to begin to shed light on structure activity relationships against WWP2 autoubiquitination. The synthesised analogues were then to be tested against other Nedd4 ligases to determine whether any degree of specificity had been acquired. The data received from this could then be used to inform further analogue development, or to select the best analogue and begin cellular based experiments. An alternative in gaining a structural view into inhibitor binding is using *in silico* methods. We computationally mapped the WWP2 HECT domain for suitable sites for binding. The site is identified by its appropriate size to accommodate small molecules and its accessibility. The putative binding site had biological relevance in terms of a mechanistic view, as it was located on the boundary of the N and C lobe, in close proximity to the hinge loop. The conformational flexibility between the N and C lobe of the HECT domain underpins its ability to transfer ubiquitin. We could therefore theorize ligands bound in this location could prevent the transfer of ubiquitin over to the

catalytic cysteine in the C lobe. The compounds identified by the Chantry lab (Jessica Watt 2018) were tested using the WWP2 FL autoubiquitination assay. Once the HECT domain was confirmed as the target domain for the compounds, we chose to focus on the inhibitor NSC288387. This compound was docked into the putative site. Inspection of the contacts around NSC288387 yielded positive results. Meaningful interactions in the form of hydrogen bonds, π - π and π -cation interactions were present. The absence of these interactions would have cast early doubt on the feasibility of this model as these intermolecular forces can be the driving force behind ligand binding events. In order to validate the *in silico* model, empirical data was needed. Ligand NMR offered the most appealing technique as epitope maps of the compounds could be built, and the newly invented DEEP-STD NMR technique can utilise the nature of the surrounding polypeptide chain to gain insight into how the ligand is binding. The epitope map of the compound matched in excellent agreement with the proposed model. The strongest contacts were observed on the *N*-phenyl ring, which was sitting deeper in the central cavity. The weakest contacts were found on the solvent exposed regions of the ethyl methoxy chain. It was clear from our NMR validated *in silico* model that the *N*-phenyl ring was sitting deepest in the binding pocket of the protein. The ligand epitope map showed the strongest contacts on that phenyl ring. Therefore, we felt any modifications to that ring had the greatest potential to form, or alter, meaningful interactions with the polypeptide chain. From a synthetic chemistry point of view, we felt modifications to that ring was the easiest way to quickly build up a small library of analogues. This was because we could first have a general route to make the tricyclic isoalloxazine core on a multi gram scale, and then derivatise using the common intermediate. The first attempted route to the isoalloxazine core started from the 2-nitroaniline. The amine underwent an S_N2 reaction with 1-bromo-2-methoxyethane, albeit in a convoluted fashion as a trifluoroacetyl group was added onto the amine first, and then removed in the S_N2 reaction. The reaction with just the primary amine and 1-bromo-2-methoxyethane did not proceed, indicating

the trifluoroacetyl group played an important part in making the reaction go to completion. The most likely reason for this was that it increases acidity and promotes alkylation. Once the alkylated nitro aniline was attained, the two step, one pot reduction and condensation reaction yielded the isoalloxazine intermediate. However, we were not satisfied with the route used, feeling the initial protection with trifluoroacetyl was unnecessary, and the S_N2 reaction was a relatively poor reaction with the production of side products and low to moderate yield. Upon re-evaluation of the synthetic strategy, it seemed obvious the most efficient approach would be to start from the 2-fluoronitrobenzene and form the desired amine via S_NAr reaction. The very nature of the fluoro *ortho* to the nitro group meant this reaction would have a high chance of going to completion. To our delight, the reaction proceeded with ease meaning we could run the reaction on a decagram scale, and using the previously described reduction and condensation, giving us access to 15 g of the key intermediate. The conditions for the *N*- aryl coupling with flavins had previously been reported, using the Chan Lam system. The first reaction with the unsubstituted phenylboronic acid gave the NSC288387 compound in moderate yield, suggesting the feasibility of this system to give us fairly rapid access to a multitude of analogues. To our disappointment, any variations to the arylboronic acid dramatically decreased the yield to under 20 %. The very low yield for some of the analogues did not present itself as a major issue at this point, as only very small quantities of compound were needed for testing and it was envisaged reaction optimization for any leading analogues would be done further down the pipeline when greater quantities would be needed. The success rate of the reaction working hovered around 50 %, meaning some of the target compounds were not synthesised. Whilst this was of some concern, we had enough compounds to begin biological testing and investigate whether there was any early indication of SAR. It is a crucial point in hit to lead optimization that the SAR is not flat. Flat SAR would indicate the compound is too difficult to optimize or the compound is not truly a non covalent inhibitor. Synthesised compounds were tested using the *in vitro* WWP2

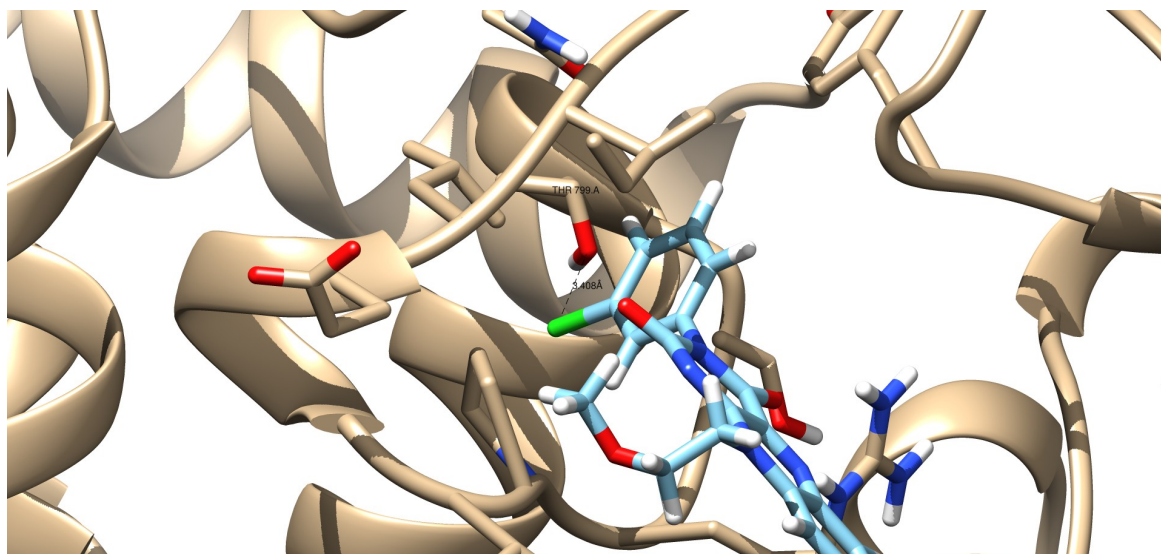


Fig. 5.22 *In silico* structure of compound **21** docked into the binding site.

autoubiquitination assay developed in our lab. The assay effectively mimics the enzymatic cascade of E1, E2 and E3, and the rate of WWP2 autoubiquitination can be quantified via an ELISA readout. The NSC288387 analogues showed significant differential alterations in inhibitory activity. Compound **21** with the *meta* chloro substituent on the N-phenyl ring showed the greatest increase in activity to 0.23 μM . Inspection of the docked model for this compound showed an additional protein ligand interaction. The chlorine atom forms a 3.4 Å halogen bond with either the carbonyl or hydroxy on Thr799, depending on the orientation. The ten fold increase in activity could possibly be attributed to the formation of this additional interaction (Fig. 5.22).

Analysis of the compounds for their inhibition against Nedd4 autoubiquitination showed a similar trend as to WWP2. Whilst there was some difference in activity of the chloro substituent, there is still fairly significant inhibition for Nedd4. The similar trend in SAR strengthens the theory of where these compounds are binding in the HECT domain. Gaining specificity towards WWP2 remains a significant task due to the high homology in amino acid sequence. However, gaining absolute specificity towards a single ubiquitin ligase may not necessarily be paramount; one can look towards the current drug targeting the ubiquitin

system. Bortezomib inhibits the proteasome, which lies at the end of the protein degradation system, is an approved drug for multiple myeloma and mantle cell lymphoma. Targeting the proteasome offers little pathway specificity and potential leads up the build up of unwanted proteins. The main issue with this treatment is the potential serious side effects a patient can experience. Despite this, it is still an efficacious treatment for cancer. As the E3 ligases confer the most specificity in the pathway, potentially only hitting other ligases in the small Nedd4 family may bring significantly less side effects than targeting the proteasome. The most obvious next step with these compounds is to start testing them in cells. Information on cell permeability and whether the compounds are hitting the target in cells is critical. One possible experiment would be to do an immunoprecipitation assay and look for levels of WWP2 substrate levels. Levels of WWP2 substrates could be analysed by western blotting and they would be expected to decrease over time, in the presence of an inhibitor it would be expected the substrate levels remain the same. Using different concentrations of compounds would determine what concentration the compounds are active in cells. These set of compounds could also be developed into PROTACs. These set of compounds would be linked with a target inhibitor such as the BET bromodomain ligand JQ1. The WWP2 inhibitors offer an alternate to the current limited number of PROTAC E3s (mainly VHL and cereblon). Alternatively, a current E3 PROTAC could target WWP2. Girardini et al reported E3 ligases hijacking each other. As our WWP2 inhibitors may not be as potent as other E3 inhibitors (μM vs nM), utilising the inhibitors as PROTACs to remove WWP2 may potentially be more effective than using the lone WWP2 inhibitors.

Chapter 6

Final Discussion

6.1 Introduction

WWP2 is part of the Nedd4 family of HECT ubiquitin ligases. The role of the highly homologous Nedd4 family of HECT ligases in disease is becoming increasingly prevalent. WWP2 targets important oncogenic pathway modulators such as Smads and Pten. Targeting WWP2 for inhibition by a small molecule could have profound therapeutic potential. There is a distinct lack of research into the development of HECT ubiquitin ligase inhibitors. Moreover, there is relatively little research into WWP2 in the context of drug discovery. This thesis aimed to first develop platforms for which small molecule inhibitors can be discovered which target WWP2. Once several inhibitors, or ligands, had been identified, elucidating the binding pocket was achieved by a combination of X-ray crystallography, data obtained from ligand NMR, molecular docking and structure activity relationship analysis. Lead compounds have emerged which have the potential for further development into a drug like compound, with possible therapeutic properties.

6.2 Gaining a starting point

When this research was started, there were no lead WWP2 inhibitors. There was also no biophysical or biochemical assays developed for WWP2 that would be suitable to detect small molecule binding. Therefore, it was important to implement a screening strategy that would be suitable for WWP2 and generate a series of compounds that could be taken further down the drug discovery pipeline. Thermal shift was the first technique chosen to screen compounds. Thermal shift offered the most appealing method as it was a simple set up and could be readily run on qPCR machines. Using the HECT domain of WWP2, over 2000 compounds were screened resulting in 11 compounds being identified as possible ligands. Validation of these compounds through STD NMR resulted in NSC35676 and NSC401005 as binders. The other identified compounds shouldn't be completely ruled out, as the NMR

results were inconclusive due to the poor quality of data. The poor quality of data may simply be due to the low aqueous solubility of the compounds. The concentration of the compounds could not be increased without increasing the DMSO concentration. Increasing the DMSO concentration caused the protein to precipitate out of solution. It was felt the confirmation of binding could be run on more sensitive techniques which required less compound. Surface plasmon resonance gave us an opportunity to validate binding, study kinetics and to develop another high throughput screening platform for WWP2. The initial experiments were promising, with already known WWP2 ligands (note: SPR experiments were chronologically done later) showing binding. However, the follow up experiments gave much more inconclusive results when attempting to run concentration dependent assays. A possible reason for this is the presence of DMSO required in the buffer. DMSO has a high refractive index, which can interfere with the SPR signal. So whilst SPR may have its benefits as a method to detect ligand binding, there is still scope for method development so it could be used for kinetic experiments or high throughput screening.

6.3 The PAIN argument

In 2010, Jonathan Baell coined the term ‘Pan Assay Interference’ compounds, also known as ‘PAINs’. PAINs are as the name suggests, promiscuous compounds which appear in multiple assays as false positives. These compounds are often reactive with elements in a screening assay. Some mechanisms by which they interfere include reactivity with any nucleophiles in a bioassay, non-specific reactivity with the protein, redox activity or there may be even be no proven mechanism. A more comprehensive description of these interference mechanisms are found in Dahlin et al. (2015). Compounds can have PAIN or close similarity moieties but still be biologically active. There are still FDA approved compounds which could visually be classed as a PAIN. During this research, we were aware that some of our hit compounds could be classed as a ‘PAIN’. We were however confident that our empirical

data from several orthogonal assays proved that the compounds were in fact true binders. Whilst working on the synthesis of the NSC288387 analogues, a particular research paper was made aware to us regarding the possibility of the compound being a PAIN. The paper in question (Raoof et al., 2013) reported the compound series toxoflavins and deazaflavins as small molecule inhibitors for a particular therapeutic target. Both are structurally similar and exhibited clear structure activity relationship. However, the toxoflavin series exhibited redox activity and produced hydrogen peroxide, proving this would be a troubling series to follow up in further experiments. What was to our concern was the structural similarities between NSC288387 and the toxoflavins. Even though we were confident that NSC288387 was a true ligand, due to the convincing NMR data, we sought to check for any possible redox activity. We set up an assay based on P.A. Johnston et al. (2011) which tests for hydrogen peroxide generated by redox cycling compounds in the presence of reducing agents (Fig. 6.1). It works through the hydrogen peroxide dependent HRP mediated oxidation of phenol red. The change in absorbance of phenol red can be measured at 610 nm. After a concentration dependent assay with NSC288387 was run, it was found the compound does produce hydrogen peroxide, but only at the higher concentrations. The concentration at which this compound was discovered as an inhibitor was at 10 μ M. Therefore a concentration dependent experiment with hydrogen peroxide was run to investigate what concentrations of H₂O₂ is needed to affect the biochemical assay. An IC₅₀ of 1.1 mM was calculated. The amount of hydrogen peroxide produced by NSC288387 at 10 μ M is negligible and not enough to interfere with the biochemical assay. It is possible if this compound or the more potent synthesised analogues are taken forward for cell based testing, that there may be some complications if the compound is slightly redox active. It could be of interest to synthesise analogues of these compounds without the isoalloxazine core. ‘Scaffold hopping’ could result in an analogous pharmacophore without the undesirable redox activity but retains the inhibitory activity.

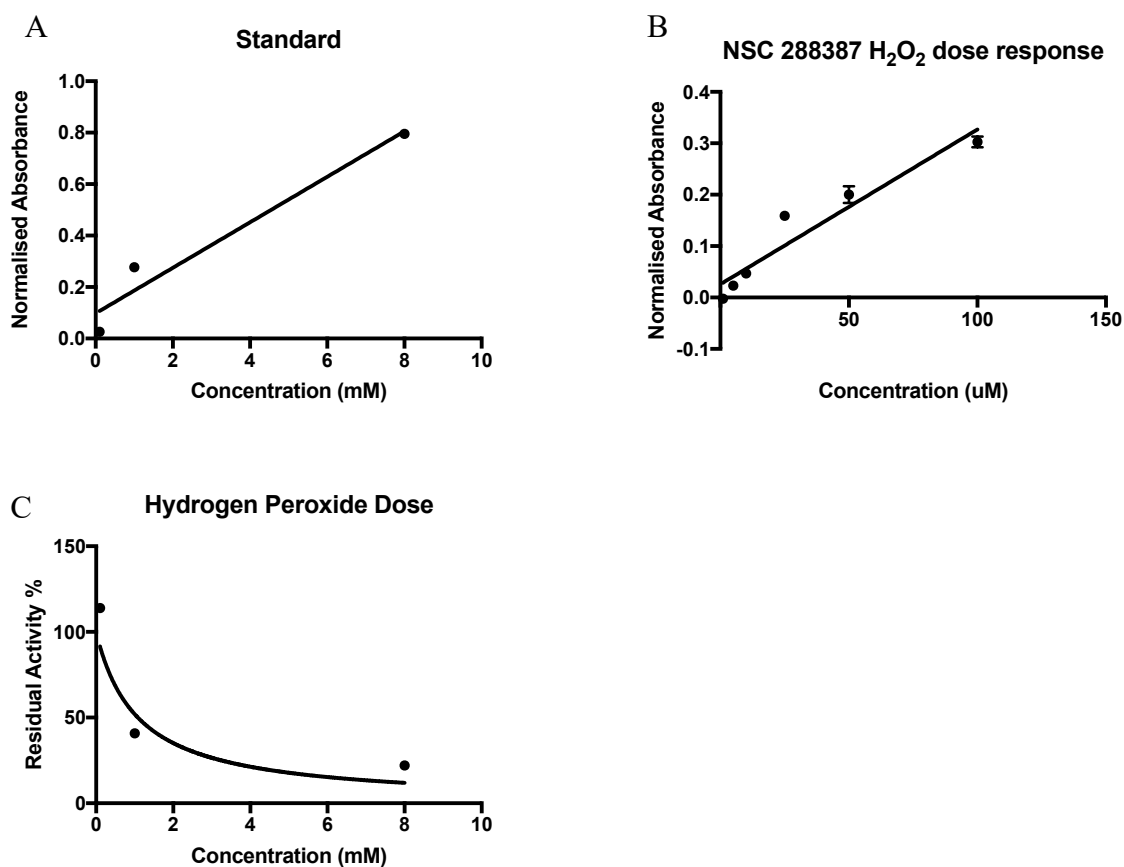


Fig. 6.1 (A) Hydrogen peroxide standard curve using the phenol red assay. (B) Phenol red assay using NSC288387. (C) Autoubiquitination concentration dependent assay using hydrogen peroxide.

Extensive soaking was attempted in order to solve the protein ligand structure. Out of all the compounds, only NSC2805 gave us a bound structure. In terms of NMR analysis, this compound was not ideal due to its symmetry. Therefore we could not validate the crystal structure using the ligand NMR techniques as we could not orientate the molecule with confidence. Using the crystal structure, unsymmetrical analogues of NSC2805 could be designed and synthesised. These analogues should contain more proton environments so meaningful epitope maps and DEEP STD data can be obtained. Combined with IC₅₀ data of these analogues the extent as to which intermolecular interactions are important can be assessed. Substantial effort went into generating a bound structure of NSC2805 to HECT

because it was the only compound which gave extra density in the binding site that was predicted in silico. A medicinal chemist may view the NSC2805 structure with scepticism. Quinones have been stated as PAIN compounds and there has been commentary on how they shouldn't be followed up as a lead compound. The argument for us to pursue a quinone compound was to establish and characterise a binding site with HECT that can induce inhibition. Whilst NSC2805 may not be "drug like", it was a useful tool in the development of methods for WWP2 inhibitor discovery.

6.4 A novel allosteric binding site

After generating some initial hits through thermal shift it was decided the generation of a structural model was imperative. The structural model would reveal how the inhibitors could possibly be built on in order to improve their potency. After harvesting over 150 crystals, 3 data sets eluded to a small molecule being bound to the HECT domain. The site where the additional electron density was picked up is where the site map predicted there could be a binding site. The electron density around NSC2805 may not be of the highest quality, but the potential protein-ligand interactions increase the likelihood that our X-ray structural models are accurate. The data in Chapter 5 strengthens the argument that the electron density is due to the ligand, and a new ligandable binding site in HECT has been discovered. This is because the docking solution of NSC288387 is in the same site as where the additional electron density was identified for NSC2805 in Chapter 4. The STD NMR derived epitope map fitted in agreement with the proposed docked solution of NSC288387. The DEEP STD data matched up with the predicted model with the expected DEEP STD factors for selected protons being observed. The differential activity of the NSC288387 analogues fits well with the docked model. For example, when the buried *N*-phenyl ring was removed, there was a drop in activity. It is the combination of this data which can give us confidence in

stating there is an allosteric binding site between the N and C lobe which can be targeted for inhibition by small molecules.

6.5 Future work

This thesis provides a framework for various avenues that could be pursued in the field of targeting the ubiquitin ligases. Fully characterising the mechanism of inhibition is an important step. Mutating key residues in the binding pocket of WWP2 HECT and measuring activity of the compounds would give an indication as to what interaction are key, and to also provide more evidence that the binding pocket is real. Improving the biophysical assays described in Chapter 3 is an obvious route. This is something we have recently started and obtained preliminary results for. After working with the protein construct explained in Chapter 2, we theorised this may be a more suitable truncation of WWP2 for thermal shift. Therefore we revisited the thermal shift assay. A test run with 2.5 μ M WW2-1-H, 5 x SYPRO orange in 25 mM Tris-Cl pH 7.4, 150 mM NaCl and 5 mM DTT was carried out. The resulting thermal shift profile was of a higher quality than the profile produced from WWP2 HECT. The slope has a steeper gradient and a more pronounced increase in fluorescence. The T_m of the protein is also higher at around 47.7 °C. The composition of the buffer is more desirable as it is at 7.4, a neutral pH which is less likely to cause reactions with certain compounds as well as being more biologically relevant. 866 compounds from the upgraded NCI diversity set VI have been screened at 10 μ M, 0.1 % DMSO. Due to time constraints the library screen was not completed.

All the developed assays stated in this thesis can be combined and run as a pipeline for WWP2 drug discovery. A more optimised compound library, perhaps from industry, which would not only be less likely to contain PAINs, but to be more structurally diverse and contain a greater number of compounds. They could be screened by thermal shift, validated by STD NMR. They could be soaked into the crystals which can now be grown routinely and soaked

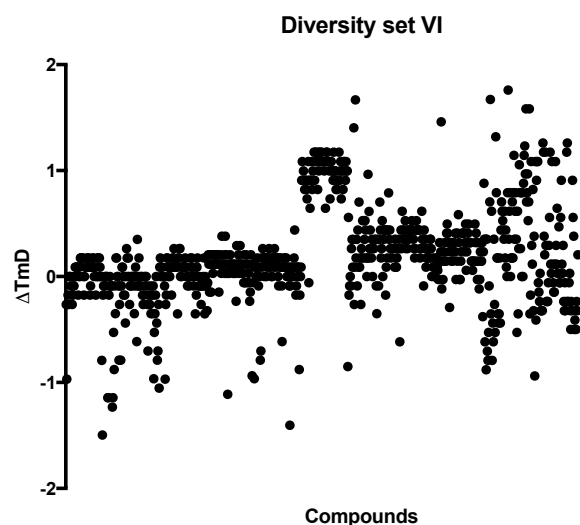


Fig. 6.2 Preliminary data showing the thermal shift screen using the WWP2 WW2-2,3 linker-HECT construct

using the ideal conditions developed in this thesis. This could lead to a ‘2nd generation’ of WWP2 inhibitors, or ligands, which are more amenable to drug discovery. Beginning the test the more potent NSC288387 analogues in cells needs to be carried out. Whether these compounds are cell permeable is an important question to ask and whether they actually inhibit WWP2 in the cell, as opposed to on a plate. The metabolic turnover of the compounds could be investigated by using mouse liver microsomes. Measuring the turnover can inform of any future organic synthesis if there is a particular moiety susceptible to rapid metabolic breakdown. A recent paper regarding WWP1 and its implication in prostate cancer brings the compound indole-3-carbinol into light (Lee et al., 2019). Indole-3-carbinol and its derivatives were described in Section 1.2.3.4 as a Nedd4 inhibitor, discovered with the help of the Nedd4 ligand structure (PDB:5C91). Kathman et al. (2015) report "compound 3" did not label WWP1 as only Nedd4 contained the non catalytic cysteine in that particular binding pocket. However, indole-3-carbinol is predicted to bind in that particular site in WWP1. A possible reason is that there are non covalent interactions which are stronger in indole-3-carbinol than in "compound 3". Due to the high homology between WWP1 and WWP2, it could be feasible that indole-3-carbinol also binds to WWP2. Soaking WWP2 crystals using the

method described in Chapter 4 could lead to another WWP2 ligand structure. This could then be used to inform synthetic modifications in order to improve specificity and potency.

Chapter 7

Chemistry experimental

7.1 General chemistry experimental

7.1.1 Preparation of solvents and glassware

All reagents obtained commercially were used without further purification, unless otherwise stated. Reactions under anhydrous conditions were carried out under a nitrogen or argon atmosphere. Anhydrous solvents were either bought in separately or dried using the following methods; tetrahydrofuran was distilled over the sodium/ benzophenone radical under an argon atmosphere, acetonitrile was stored overnight under argon over activated 3 Å molecular sieves, dichloromethane was distilled over calcium hydride and *N,N*-dimethylformamide was dried overnight under argon over 4 Å molecular sieves, vacuum distilled and stored under argon.

7.1.2 Compound characterisation

^1H , ^{13}C and ^{19}F NMR spectra were gathered using either a Bruker Ascend 500 at 500 MHz, 126 MHz and 471 MHz or a Bruker Ultrashield 400 Plus at 400 MHz, 100 MHz and 367 MHz. Chemical shifts are reported in ppm and coupling constants (*J* values) reported in Hz. HSQC and COSY experiments were run for novel compounds. High resolution mass spectrometry data was collected from the EPSRC UK National Mass Spectrometry service at the University of Swansea.

Liquid chromatography mass spectroscopy (LCMS) was carried out using a Water Acquity I Class Ultra Performance Liquid Chromatography Mass Spectrometry machine fitted with a Waters Acquity UPLC BEH C18 column (2.1 x 50mm 1.7 μm). Samples were dissolved in 2 mL acetonitrile: water (50:50). Samples were filtered through a 0.45 μm filter before use. All LCMS experiments were run at Charnwood Molecular at Biocity, Nottingham. Infrared spectra were obtained using a Perkin Elmer Spectrum 100 Fourier

Transformation Infrared Spectrometer. Samples were ran as a thin film from an evaporated solution of dichloromethane on sodium chloride plates. Melting points were carried out on a Buchi Melting Point B-545 apparatus.

7.1.3 Chromatographic techniques

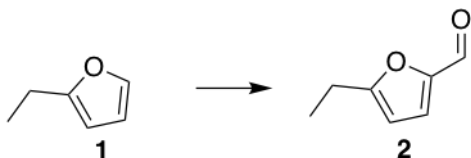
Flash column chromatography was carried out either using conventional glass columns or using a Biotage IsoleraTM Prime automated flash purification system. Flash chromatography using glass columns were manually packed with Material Harvest silica gel 60. Flash chromatography using the Biotage IsoleraTM Prime system was carried out using 5- 120 g Biotage ZIP cartridges, packed with Biotage KP-Sil brand silica. Column chromatography using the Biotage system was carried out at Charnwood Molecular in Biocity, Nottingham. Thin layer chromatography was carried out on Merk aluminium plates coated with Kieselgel 60 F254 silica gel. Plates were visualised under UV light at 254 nm (short wave) or 365 nm (long wave).

7.1.4 Compound numbering

All compounds involved in synthetic procedures are numbered **1-33** in bold. All other compounds referred to in this thesis, not involved in the synthetic chemistry, are labelled with their reference name/ number e.g. NSC35676. The exceptions are compound **18**, which is also referred to as NSC288387, and compound **7** which is also referred to as heclin. "Compound 3" and "Compound 1" are in reference to the inhibitors reported by Kathman et al.

7.2 Individual procedures and characterisation

5-Ethylfuran 2-carbaldehyde (2)



To a mixture of DMF (1 mL), was added phosphorous oxychloride (1.26 mL, 13.5 mmol, 1.35 equiv) and was stirred at 0 °C for 20 minutes. 2-Ethylfuran (1 g, 10.0 mmol) was added at a rate that the reaction did not rise above 20 °C. After the addition was complete, the reaction mixture was kept at 0-5 °C for 1 hour followed by 1 hour at room temperature. The mixture was poured into approximately 20 mL of ice and water. The mixture was neutralised with sodium carbonate and was left to stand overnight. Diethyl ether was added and the organic layer was separated. The aqueous layer was extracted three times with ether. The organic extracts were combined and dried over sodium sulfate. The solvent was removed under reduced pressure. The crude was purified by fractional distillation to afford a brown oil. (1.1 g, 79 %).

^1H NMR (500 MHz, CDCl_3) δ 9.53 (s, 1H), 7.18 (d, $J = 3.5$ Hz, 1H), 6.24 (d, $J = 3.5$ Hz, 1H), 2.76 (q, $J = 7.3$ Hz, 2H), 1.30 (t, $J = 7.6$ Hz, 3H).

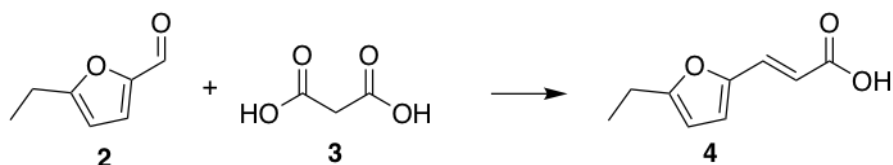
^{13}C NMR (126 MHz, DMSO) δ 209.9, 174.3, 163.5, 36.8, 35.3, 28.2, 8.1.

IR ν_{max} (film)/ cm^{-1} : 1713, 1116

Procdure- (Taylor, 1959)

Data- (Kuo and Shih, 1991)

3-(5-Ethyl-2-furanyl)-2-propenoic acid (**4**)



Piperidine (0.8 mL, 8 mmol, 1 equiv.) was added to a solution of **2** (1 g, 8 mmol), malonic acid (0.62 mL, 10 mmol, 1.25 equiv) and pyridine (1.42 mL, 17 mmol, 2.13 equiv) under nitrogen atmosphere. The reaction mixture was heated to 130 °C under vigorous stirring for 3 hours. The reaction mixture was cooled and ethyl acetate was added. The mixture was acidified to pH 5 with 1 M hydrochloric acid and the organic layer was separated. The organic layer was washed with water and dried over magnesium sulfate. The solvent was removed to afford a yellow solid (630 mg, 47%).

¹H NMR (500 MHz, CDCl₃) δ 7.45 (d, J = 15.6 Hz, 1H), 6.58 (d, J = 3.3 Hz, 1H), 6.23 (d, J = 15.6 Hz, 1H), 6.11 (dt, J = 3.3, 0.9 Hz, 1H), 2.70 (q, J = 7.6 Hz, 2H), 1.27 (t, J = 7.6 Hz, 3H).

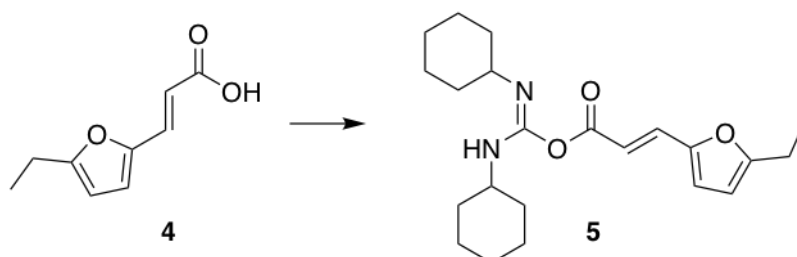
¹³C NMR (126 MHz, CDCl₃) δ 172.1, 161.8, 149.2, 133.2, 117.4, 112.8, 107.5, 21.7, 11.9.

IR ν_{max} (film)/cm⁻¹: 3425, 1629.

Mp 127-128 °C

Procedure and data- (Van Veldhoven et al., 2011)

(Z)-N,N'-dicyclohexylcarbamiimidic (E)-3-(5-ethylfuran-2-yl)acrylic anhydride (5)



Compound **4** (270 mg, 1.6 mmol) *N, N'*-dicyclohexylcarbodiimide (DCC) (400mg, 2.0 mmol, 1.25 equiv) and 4-dimethylaminopyridine (DMAP) (30 mg, 0.25 mmol, 15.6 %) were dissolved in anhydrous dichloromethane (10 mL). The solution was cooled to 0 °C and stirred for 30 minutes. A solution of 4-acetylaniline (200 mg, 1.5 mmol, 0.94 equiv) in DCM was added dropwise over 30 minutes. The mixture was warmed to room temperature and stirred overnight. *N, N'*-dicyclohexylcarbodiimide was filtered off and the solvent was removed under reduced pressure. The crude product was purified by column chromatography (3:1 petrol: ethyl acetate) to give **5** as a waxy colourless solid (220 mg, 39 %).

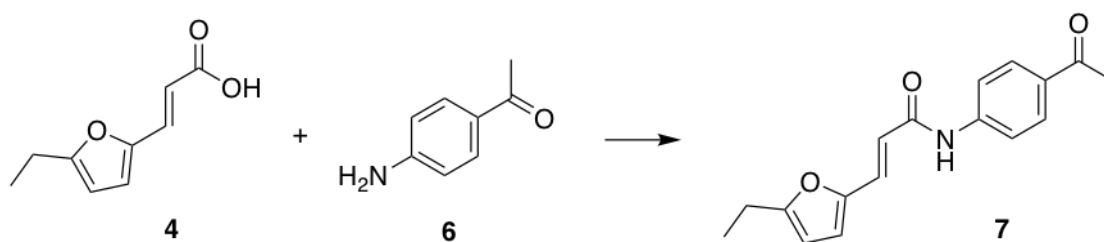
¹H NMR (500 MHz, CDCl₃) δ 7.37 (d, *J* = 15.0 Hz, 1H), 6.57 (d, *J* = 15.0 Hz, 1H), 6.49 (d, *J* = 3.1 Hz, 1H), 6.07 (dd, *J* = 3.1, 0.9 Hz, 1H), 2.74 – 2.60 (m, 2H), 2.04 – 1.92 (m, 4H), 1.82 (m, 15.0 Hz, 4H), 1.78 – 1.69 (m, 2H), 1.46 – 1.11 (m, 12H).

¹³C NMR (126 MHz, CDCl₃) δ 167.20, 160.78, 154.26, 149.85, 130.28, 116.24, 115.28, 107.29, 56.22, 49.79, 32.74, 30.93, 26.34, 25.53, 25.41, 24.71, 21.67, 11.87.

IR ν_{max} (film)/cm⁻¹: 3293, 3051, 2933, 2855, 1701, 1527

m/z (NSI-FTMS) [M+H]⁺: calcd for [C₂₂H₃₃N₂O₃]⁺ 373.2486, found 373.2486

***N*-(4-Acetylphenyl)-3-(5-ethyl-2-furanyl)-2-propenamide (5)**



Compound **4** (274 mg, 1.7 mmol, 1.1 equiv), DCC (400 mg, 2.0 mmol, 1.3 equiv) and DMAP (30 mg, 0.25 mmol, 16.6 %) were dissolved in anhydrous dichloromethane (10 mL). The solution was cooled to 0 °C and stirred for 30 minutes. A solution of 4-acetylaniline (200mg, 1.5 mmol) in DCM was added dropwise over 30 minutes. The mixture was refluxed at 50 °C and stirred overnight. *N,N'*-dicyclohexylcarbodiimide was filtered off and the solvent was removed under reduced pressure. The crude product was purified by column chromatography (3:1 petrol: ethyl acetate) to give a yellow solid (110 mg, 26 %).

¹H NMR (500 MHz, CDCl₃) δ 7.97 – 7.94 (m, 1H), 7.74 – 7.67 (m, 2H), 7.48 (d, *J* = 15.0 Hz, 1H), 6.55 (d, *J* = 3.3 Hz, 1H), 6.38 (d, *J* = 15.0 Hz, 1H), 6.10 (dt, *J* = 3.3, 0.9 Hz, 1H), 2.72 – 2.66 (q, *J* = 7.6 Hz, 2H), 2.58 (s, 3H), 1.29 – 1.25 (t, *J* = 7.6 Hz, 3H).

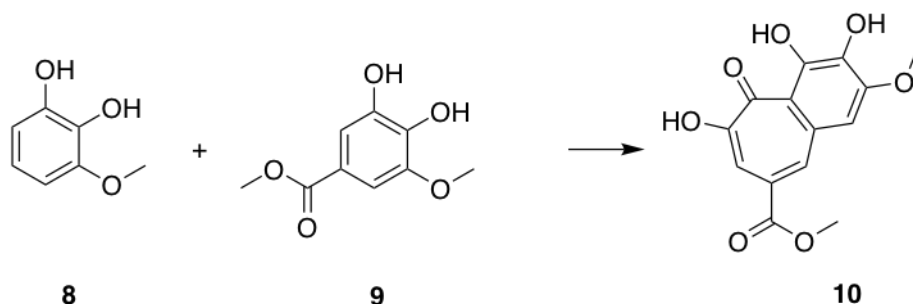
¹³C NMR (126 MHz, CDCl₃) δ 196.9, 164.4, 161.0, 149.5, 142.6, 132.8, 130.1, 129.8, 118.9, 116.8, 115.9, 107.5, 26.5, 21.7, 11.9.

IR ν_{max} (film)/cm⁻¹: 3432, 1654

Mp 130-132 °C

Procedure- (Arias et al., 2011)

Methyl 3,4,6-trihydroxy-2-methoxy-5-oxo-5H-benzo[7]annulene-8-carboxylate (10)



3-Methoxy catechol (140 mg, 1 mmol) and methyl 3, 4, 5 trihydroxy benzoate (185 mg, 1 mmol, 1 equiv) was dissolved in a mixture of acetone phosphate- citrate (1:1 0.2M Na₂HPO₄: 0.1 M citrate) buffer (1:5 v/v) containing 1 mg horseradish peroxidase. Hydrogen peroxide (3%) was added in 2 mL portions every 10 minutes over a 40 minute period. The precipitate was filtered off and washed with 3 x 6 mL water. The solid was dried under vacuum and purified by hot filtration in ethanol to give an orange solid (146 mg, 50 %.)

¹H NMR (500 MHz, DMSO) δ 8.39 (d, J = 1.5 Hz, 1H), 7.58 (d, J = 1.6 Hz, 1H), 7.47 (s, 1H), 4.01 (s, 3H), 3.90 (s, 3H).

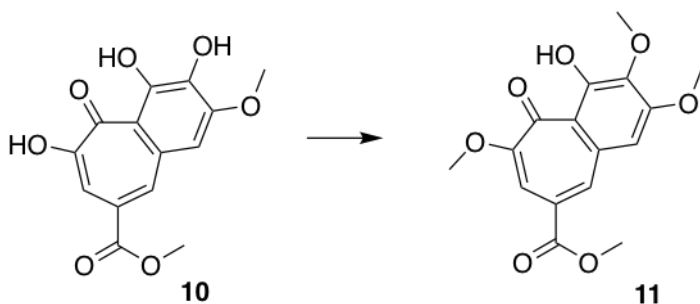
¹³C NMR (126 MHz, DMSO) δ 183.6, 167.1, 154.2, 152.5, 151.9, 138.5, 138.3, 130.2, 123.9, 116.7, 114.5, 110.5, 56.7, 53.4.

IR ν_{max} (film)/cm⁻¹: 3428, 1639.

Mp 215-217 °C decomposition.

Procedure and data- (Cheng et al., 2012)

Methyl 4-hydroxy-2,3,6-trimethoxy-5-oxo-5H-benzo[7]annulene-8-carboxylate (11)



To a suspension of **10** (200 mg, 0.68 mmol, 0.33 equiv), K_2CO_3 (188 mg, 1.36 mmol, 0.66 equiv) in anhydrous DMF (35 mL) was added iodomethane (0.31 mL, 2.05 mmol) dropwise at 0 °C under argon. The mixture was stirred overnight at room temperature. The reaction was quenched with water and the mixture was stirred for 20 minutes. EtOAc was added and the organic was washed with water and the combined aqueous layers were extracted with EtOAc. The organic was dried over Na_2SO_4 and the solvent was removed under reduced pressure. The crude was purified using column chromatography (2:1 petrol:ethyl acetate) to yield a yellow solid (170 mg, 64 %).

1H NMR (500 MHz, $CDCl_3$) δ 14.93 (s, 1H), 8.20 (d, J = 0.6 Hz, 1H), 7.34 (d, J = 0.5 Hz, 1H), 6.87 (s, 1H), 4.03 (s, 3H), 4.01 (s, 3H), 3.98 (s, 3H), 3.97 (s, 3H).

^{13}C NMR (126 MHz, DMSO) δ 184.9, 166.6, 158.1, 156.5, 156.2, 138.2, 137.8, 132.9, 124.1, 117.2, 110.0, 109.6, 56.3, 56.1, 53.1

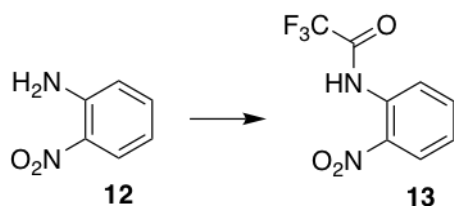
IR ν_{max} (film)/ cm^{-1} : 3434, 2081, 1638

m/z (TOF MS) $[M+H]^+$: calcd for $[C_{16}H_{16}O_7]^+$ 321.0974, found 321.0971.

Mp 162-164 °C

Procedure- (Cheng et al., 2012)

2,2,2-Trifluoro-*N*-(2-nitrophenyl)acetamide



2-Nitro aniline (4.14 g, 30 mmol) was dissolved in dichloromethane (60 mL) and was cooled to 0 °C and stirred for 1 hour. Trifluoroacetic anhydride (8.4 mL, 60 mmol) was added and the reaction mixture was stirred for another 1 hour. NaHCO₃ was added to neutralise the reaction. The organic layer was separated and dried over Na₂SO₄. The solvent was removed to afford a light brown solid (6.36 g, 91 %). No further purification is required.

¹H NMR (400 MHz, CDCl₃) δ 11.52 – 11.13 (m, 1H), 8.91 – 8.63 (m, 1H), 8.31 (dd, J = 8.5, 1.5 Hz, 1H), 7.76 (ddd, J = 8.6, 7.5, 1.3 Hz, 1H), 7.37 (ddd, J = 8.6, 7.4, 1.3 Hz, 1H).

¹³C NMR (101 MHz, CDCl₃) δ 155.36 (q, J^2 CF = 38.4 Hz), 137.0, 136.4, 132.1, 125.6, 126.2, 122.2, 115.4 (q, J^1 CF = 288.7 Hz).

¹⁹F NMR (376 MHz, CDCl₃) δ -76.24 (d, J = 1.2 Hz).

IR ν_{max} (film)/cm⁻¹: 3434, 2094, 1638, 1729, 1172.

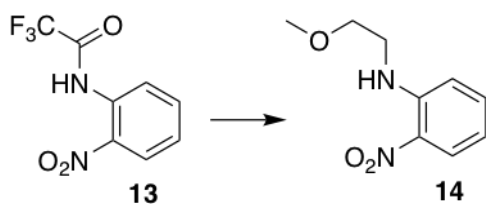
LCMS m/z (ESI) [M+H]⁺: calcd for [C₈H₅F₃N₂O₃]⁺ 233.03, found 233.1

Mp 89-91 °C

Data- (Darvesh et al., 2006)

Procedure- (Epple et al., 1997)

***N*-(2-methoxyethyl)-2-nitroaniline (14)**



13 (3 g, 12 mmol) and K_2CO_3 (6.3 g, 46 mmol, 3.6 equiv) in DMF (40 mL) was heated to 100 °C. The reaction mixture was stirred for 1 hour. 1- Bromo-2-methoxyethane (3.16 mL, 33.6 mmol, 2.8 equiv) was added and stirred overnight. The reaction mixture was cooled and H_2O (200 mL) was added. The product was extracted with DCM (3x 200 mL) and washed with brine (3 x 200 mL) The organic was dried over Na_2SO_4 and the solvent was removed to afford an orange oil. The crude was purified by flash chromatography (100 % DCM) to afford the product as an orange oil (640 mg, 56 %)

^1H NMR (500 MHz, CDCl_3) δ 8.24 – 8.13 (m, 1H), 7.49 – 7.41 (m, 1H), 6.92 – 6.83 (m, 1H), 6.66 (ddd, $J = 8.4, 6.9, 1.2$ Hz, 1H), 3.68 (t, $J = 5.4$ Hz, 2H), 3.49 (t, $J = 5.4$ Hz, 2H), 3.43 (s, 3H).

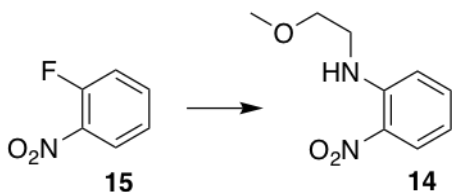
^{13}C NMR (126 MHz, CDCl_3) δ 136.2, 126.9, 115.5, 113.8, 70.5, 59.1, 42.8.

IR ν_{max} (film)/ cm^{-1} : 3379, 1619, 1571, 1419, 1352, 1262, 1155, 1120,

LCMS m/z (ESI) $[\text{M}+\text{H}]^+$: calcd for $[\text{C}_9\text{H}_{12}\text{N}_2\text{O}_3]^+$ 197.08, found 197.1

Procedure- (Epple et al., 1997)

***N*-(2-methoxyethyl)-2-nitroaniline (14)**



A mixture of 2-fluoronitrobenzene (20 g, 142 mmol) and sodium carbonate (7.525 g, 71 mmol, 0.5 equiv) in ethanol (57 mL) was heated to 80 °C. 2- Methoxyethylamine (0.67 mL, 7.7 mmol, 0.05 equiv) was added dropwise and the reaction was left overnight. The solvent was removed under reduced pressure and water was added. The product was extracted with DCM and dried over Na₂SO₄. The solvent was removed to afford an orange oil (20 g, 72 %).

¹H NMR (500 MHz, CDCl₃) δ 8.24 – 8.13 (m, 1H), 7.49 – 7.41 (m, 1H), 6.92 – 6.83 (m, 1H), 6.66 (ddd, *J* = 8.4, 6.9, 1.2 Hz, 1H), 3.68 (t, *J* = 5.4 Hz, 2H), 3.49 (t, *J* = 5.4 Hz, 2H), 3.43 (s, 3H).

¹³C NMR (126 MHz, CDCl₃) δ 136.1, 126.9, 115.5, 113.8, 70.5, 59.1, 42.8.

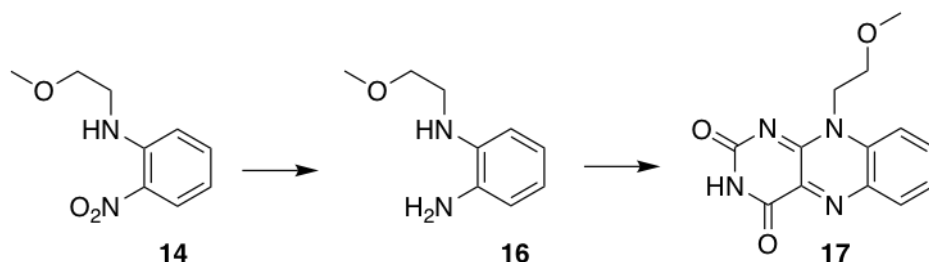
IR ν_{max} (film)/cm⁻¹: 3379, 1619, 1571, 1419, 1352, 1262, 1155, 1120,

LCMS *m/z* (ESI) [M+H]⁺: calcd for [C₉H₁₂N₂O₃]⁺ 197.08, found 197.1

Method- (Yao et al., 2017)

Unable to access published spectra

10-(2-Methoxyethyl)benzo[g]pteridine-2,4(3H,10H)-dione (17)



Compound **14** (20 g, 102 mmol) was added to acetic acid (408 mL). The mixture was degassed and Pd/C (325 mg, 3.06 mmol, 3 %) was added. The mixture was further degassed. The flask was placed under vacuum and purged with hydrogen. The reaction was stirred at room temperature overnight. The slurry was filtered through celite and alloxan monohydrate (32.66 g, 204 mmol, 2 equiv) and boric acid (31.53 g, 510 mmol, 5 equiv) was added and the mixture was stirred overnight at room temperature. The acetic acid was removed under vacuum and water was added. The reaction mixture was extracted with DCM and the organic was dried over Na₂SO₄. The solvent was removed and the resulting solid was found to be majority starting material by LCMS. The aqueous layer was filtered and the collected solid was washed with DCM to afford a green solid. (14.791 g, 53 %).

¹H NMR (500 MHz, CDCl₃) δ 8.45 (s, 1H), 8.31 (dd, J = 8.2, 1.3 Hz, 1H), 7.91 (ddd, J = 14.5, 10.3, 4.8 Hz, 2H), 7.76 – 7.62 (m, 1H), 4.93 (t, J = 5.1 Hz, 2H), 3.94 (t, J = 5.1 Hz, 2H), 3.29 (s, 3H).

¹³C NMR (126 MHz, DMSO) δ 160.2, 156.1, 151.1, 139.1, 135.2, 133.6, 132.0, 126.5, 117.5, 68.9, 58.9, 44.6.

IR ν_{max} (film)/cm⁻¹: 3418, 1670, 1538, 1508

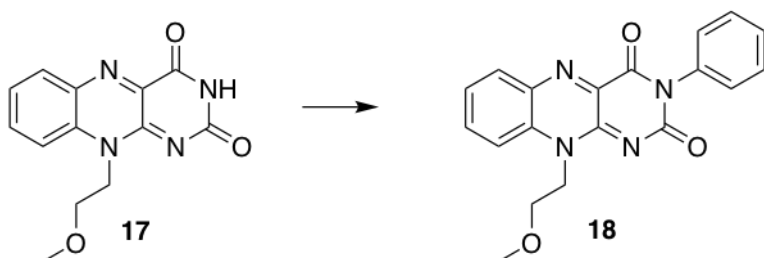
LCMS m/z (ESI) [M+H]⁺: calcd for [C₁₃H₁₂N₄O₃]⁺ 273.09, found 273.1

Mp >220 °C

Procedure- (Epple et al., 1997)

Compound in literature (unable to access)- (Lacroix et al., 1987)

10-(2-Methoxyethyl)-3-phenylbenzo[g]pteridine-2,4(3H,10H)-dione (18)



To a solution of **17** (22 mg, 0.08 mmol) in anhydrous DMF (1 mL) was added benzene boronic acid (30 mg, 0.24 mmol, 3 equiv), Cu(OAc)₂ (30 mg, 0.16 mmol, 2 equiv), triethylamine (0.03 mL, 0.24 mmol, 3 equiv) and was stirred overnight at room temperature in the dark. The reaction was monitored by LCMS and once gone to completion, the mixture was diluted with DCM and the organic was washed with water and brine. The organic was dried over Na₂SO₄ and the solvent was removed. The crude solid was purified by flash chromatography (100 % EtOAc) to afford a yellow solid (27 mg, 57 %).

¹H NMR (500 MHz, CDCl₃) δ 8.32 (d, J = 7.7 Hz, 1H), 7.95 (d, J = 8.6 Hz, 1H), 7.92 – 7.84 (m, 1H), 7.63 (t, J = 7.4 Hz, 1H), 7.53 (t, J = 7.7 Hz, 2H), 7.46 (t, J = 7.5 Hz, 1H), 7.32 (d, J = 7.6 Hz, 2H), 4.96 (t, J = 4.9 Hz, 2H), 3.97 (t, J = 5.1 Hz, 2H), 3.31 (s, 3H).

¹³C NMR (101 MHz, CDCl₃) δ 155.2, 136.0, 135.6, 135.5, 132.9, 129.4, 128.7, 128.2, 126.6, 116.9, 69.6, 59.2, 45.7.

IR ν_{max} (film)/cm⁻¹: 3412, 1666, 1556

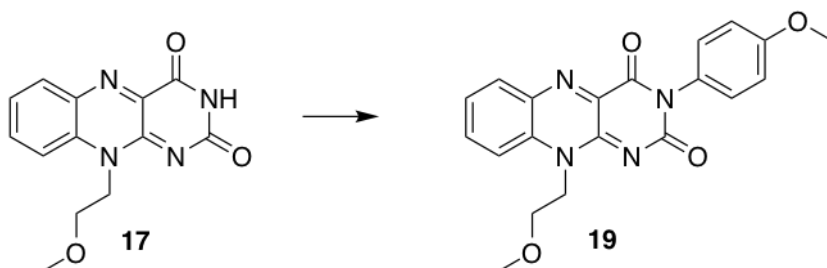
LCMS m/z (ESI) [M+H]⁺: calcd for [C₁₉H₁₆N₄O₃]⁺ 349.12, found 349.1

Mp >220 °C

Procedure- (Attenberger et al., 2008)

Compound in literature (unable to access)- (Lacroix et al., 1987)

10-(2-Methoxyethyl)-3-(4-methoxyphenyl)benzo[g]pteridine-2,4(3H,10H)-dione (19)



To a solution of **17** (272 mg, 1 mmol) in anhydrous DMF (12.5 mL) was added 4-methoxybenzene boronic acid (456 mg, 3 mmol, 3 equiv), Cu(OAc)₂ (363 mg, 2 mmol, 2 equiv), triethylamine (0.42 mL, 3 mmol, 3 equiv) and was stirred overnight at room temperature in the dark. The reaction was monitored by LCMS and once gone to completion, the mixture was diluted with CH₂Cl₂ and the organic was washed with water and brine. The organic was dried over Na₂SO₄ and the solvent was removed. The crude solid was purified by flash chromatography (9:1 CH₂Cl₂:MeOH) to afford a yellow solid (72 mg, 19%).

¹H NMR (400 MHz, CDCl₃) δ 8.31 (d, *J* = 7.5 Hz, 1H), 8.01 – 7.81 (m, 2H), 7.62 (t, *J* = 7.1 Hz, 1H), 7.25 – 7.18 (m, 1H), 7.09 – 7.00 (m, 2H), 6.79 – 6.72 (m, 1H), 4.96 (t, *J* = 5.1 Hz, 2H), 3.96 (t, *J* = 5.1 Hz, 2H), 3.86 (s, 3H), 3.31 (s, 3H).

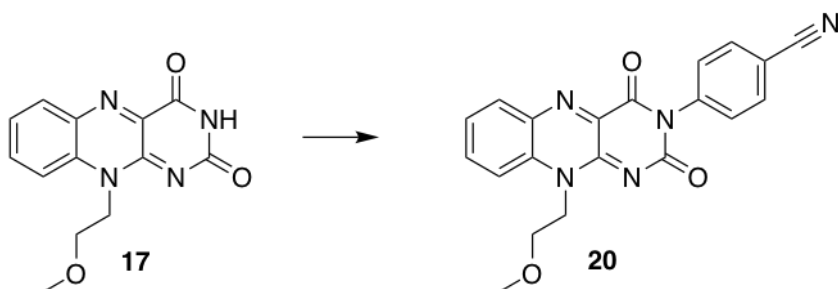
¹³C NMR (126 MHz, CDCl₃) δ 159.5, 155.6, 153.5, 149.9, 136.0, 135.6, 132.9, 129.1, 128.0, 126.5, 116.9, 116.1, 114.7, 114.7, 69.6, 59.2, 55.8, 55.5, 45.7.

IR ν_{max} (film)/cm⁻¹: 2934, 1667, 1558, 1510, 1247

LCMS *m/z* (ESI) [M+H]⁺: calcd for [C₂₀H₁₈N₄O₂]⁺ 379.13, found 379.2

Mp >220 °C

4-(10-(2-Methoxyethyl)-2,4-dioxobenzo[g]pteridin-3(2H,4H,10H)-yl)benzonitrile (20)



To a solution of **17** (272 mg, 1 mmol) in anhydrous DMF (12.5 mL) was added 4-cyanobenzene boronic acid (440 mg, 3 mmol, 3 equiv), Cu(OAc)₂ (363 mg, 2 mmol, 2 equiv), triethylamine (0.42 mL, 3 mmol, 3 equiv) and was stirred overnight at room temperature in the dark. The reaction was monitored by LCMS and once gone to completion, the mixture was diluted with CH₂Cl₂ and the organic was washed with water and brine. The organic was dried over Na₂SO₄ and the solvent was removed. The crude solid was purified by flash chromatography (100 % EtOAc) to afford a yellow solid (24 mg, 6 %).

¹H NMR (500 MHz, CDCl₃) δ 8.34 (s, 1H), 8.00 (d, *J* = 8.3 Hz, 1H), 7.96 – 7.90 (m, 1H), 7.81 (d, *J* = 8.4 Hz, 2H), 7.69 (dd, *J* = 13.2, 4.8 Hz, 1H), 7.52 – 7.41 (m, 2H), 4.97 (t, *J* = 4.7 Hz, 2H), 3.96 (t, *J* = 4.9 Hz, 2H), 3.30 (s, 3H).

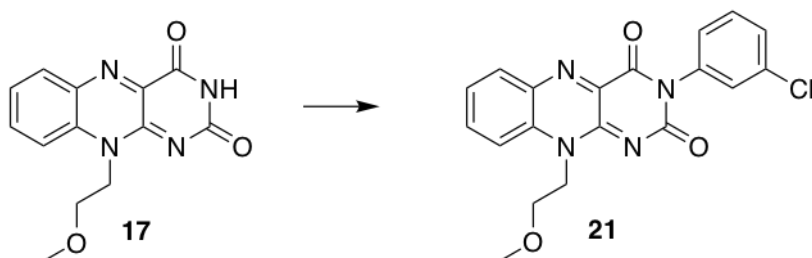
¹³C NMR (126 MHz, CDCl₃) δ 154.5, 139.5, 136.1, 133.2, 133.1, 129.6, 126.9, 117.2, 112.7, 69.5, 59.3, 46.0.

IR ν_{max} (film)/cm⁻¹: 3427, 2228, 1716, 1668, 1557

LCMS *m/z* (ESI) [M+H]⁺: calcd for [C₂₀H₁₅N₅O₃]⁺ 374.12, found 374.2

Mp >220 °C

3-(3-Chlorophenyl)-10-(2-methoxyethyl)benzo[g]pteridine-2,4(3H,10H)-dione (21)



To a solution of **17** (272 mg, 1 mmol) in anhydrous DMF (12.5 mL) was added 3-chlorobenzene boronic acid (469 mg, 3 mmol, 3 equiv), Cu(OAc)₂ (363 mg, 2 mmol, 2 equiv), triethylamine (0.42 mL, 3 mmol, 3 equiv) and was stirred overnight at room temperature in the dark. The reaction was monitored by LCMS and once gone to completion, the mixture was diluted with CH₂Cl₂ and the organic was washed with water and brine. The organic was dried over Na₂SO₄ and the solvent was removed. The crude solid was purified by flash chromatography (100% EtOAc) to afford a yellow solid (16 mg, 4 %).

¹H NMR (500 MHz, CDCl₃) δ 8.32 (dd, J = 8.2, 1.3 Hz, 1H), 7.96 (d, J = 8.1 Hz, 1H), 7.90 (ddd, J = 8.7, 7.1, 1.5 Hz, 1H), 7.69 – 7.61 (m, 1H), 7.48 – 7.41 (m, 2H), 7.33 (t, J = 1.8 Hz, 1H), 7.22 (dt, J = 7.4, 1.7 Hz, 1H), 4.96 (t, J = 5.0 Hz, 2H), 3.97 (t, J = 5.1 Hz, 2H), 3.31 (s, 3H).

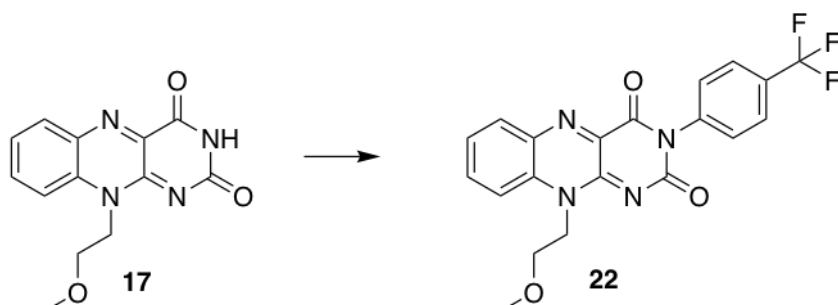
¹³C NMR (126 MHz, CDCl₃) δ 135.83, 134.86, 133.05, 130.32, 129.09, 128.78, 126.75, 126.67, 117.00, 69.57, 59.27, 45.85.

IR ν_{max} (film)/cm⁻¹: 1716, 1669, 1557, 765

LCMS m/z (ESI) [M+H]⁺: calcd for [C₁₉H₁₅ClN₄O₃]⁺ 383.08, found 383.1

Mp >220 °C

10-(2-Methoxyethyl)-3-(4-(trifluoromethyl)phenyl)benzo[g]pteridine-2,4(3H,10H)-dione
(22)



To a solution of **17** (136 mg, 0.5 mmol) in anhydrous DMF (6.25 mL) was added 4-trifluoromethylbenzene boronic acid (286 mg, 1.5 mmol, 3 equiv), Cu(OAc)₂ (181 mg, 1 mmol, 2 equiv), triethylamine (0.21 mL, 1.5 mmol, 3 equiv) and was stirred overnight at room temperature in the dark. The reaction was monitored by LCMS and once gone to completion, the mixture was diluted with CH₂Cl₂ and the organic was washed with water and brine. The organic was dried over Na₂SO₄ and the solvent was removed. The crude solid was purified by flash chromatography (100 % EtOAc) to afford a yellow solid (11 mg, 5 %).

¹H NMR (500 MHz, CDCl₃) δ 8.33 (d, *J* = 8.2 Hz, 1H), 7.94 (dt, *J* = 15.8, 8.7 Hz, 2H), 7.80 (d, *J* = 8.3 Hz, 2H), 7.66 (t, *J* = 7.3 Hz, 1H), 7.55 – 7.41 (m, 2H), 4.98 (t, *J* = 5.0 Hz, 2H), 3.98 (t, *J* = 5.1 Hz, 2H), 3.31 (s, 3H).

¹³C NMR (126 MHz, CDCl₃) δ 154.44, 139.53, 136.11, 133.25, 133.08, 129.59, 126.96 (q, *J* CF = 3.7 Hz), 117.15, 112.70, 69.52, 59.27, 46.01.

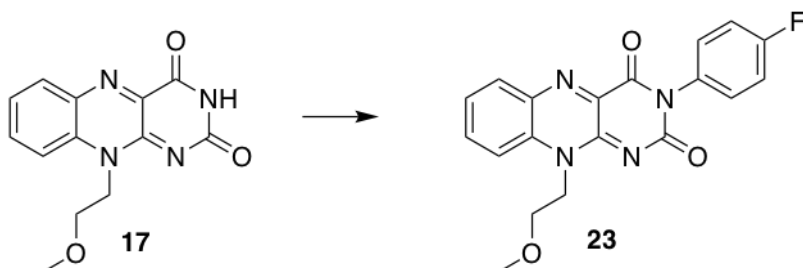
¹⁹F NMR (376 MHz, CDCl₃) δ -62.64.

IR ν_{max} (film)/cm⁻¹: 3422, 1663, 1558, 1323, 1063

LCMS *m/z* (ESI) [M+H]⁺: calcd for [C₁₉H₁₆N₄O₃]⁺ 349.12, found 349.1

Mp >220 °C

3-(4-Fluorophenyl)-10-(2-methoxyethyl)benzo[g]pteridine-2,4(3H,10H)-dione (23)



To a solution of **17** (136 mg, 0.5 mmol) in anhydrous DMF (12.5 mL) was added 4-fluorobenzene boronic acid (420 mg, 3 mmol, 6 equiv), Cu(OAc)₂ (363 mg, 2 mmol, 4 equiv), triethylamine (0.42 mL, 3 mmol, 6 equiv) and was stirred overnight at room temperature in the dark. The reaction was monitored by LCMS and once gone to completion, the mixture was diluted with CH₂Cl₂ and the organic was washed with water and brine. The organic was dried over Na₂SO₄ and the solvent was removed. The crude solid was purified by flash chromatography (100 % EtOAc) to afford a yellow solid (63 mg, 17 %).

¹H NMR (500 MHz, CDCl₃) δ 8.32 (dd, *J* = 8.2, 1.3 Hz, 1H), 7.96 (d, *J* = 8.2 Hz, 1H), 7.90 (ddd, *J* = 8.7, 7.1, 1.5 Hz, 1H), 7.65 (dd, *J* = 11.2, 4.1 Hz, 1H), 7.29 (ddd, *J* = 10.0, 5.0, 2.6 Hz, 2H), 7.21 (ddd, *J* = 12.1, 6.1, 2.7 Hz, 2H), 4.96 (t, *J* = 5.1 Hz, 2H), 3.97 (t, *J* = 5.1 Hz, 2H), 3.31 (s, 3H).

¹³C NMR (101 MHz, CDCl₃) δ 155.1, 136.1, 135.8, 133.0, 130.0, 129.9, 126.7, 116.9, 116.6, 116.3, 69.6, 59.2, 45.8. *CF coupling was not observed in the spectra.

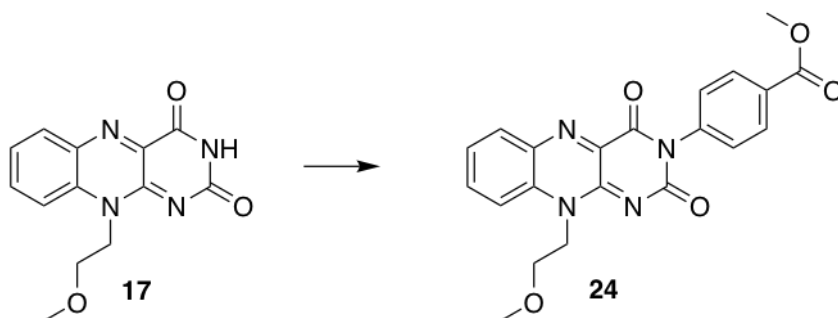
¹⁹F NMR (376 MHz, CDCl₃) δ -112.97.

IR ν_{max} (film)/cm⁻¹: 3426, 2086, 1638

LCMS *m/z* (ESI) [M+H]⁺: calcd for [C₂₀H₁₅F₃N₄O₃]⁺ 367.11, found 367.2

Mp >220 °C

Methyl 4-(10-(2-methoxyethyl)-2,4-dioxobenzo[g]pteridin-3(2H,4H,10H)-yl)benzoate



To a solution of **17** (272 mg, 1 mmol) in anhydrous DMF (12.5 mL) was added 4-methoxycarbonylphenyl boronic acid (540 mg, 3 mmol, 3 equiv), Cu(OAc)₂ (363 mg, 2 mmol, 2 equiv), triethylamine (0.42 mL, 3 mmol, 3 equiv) and was stirred overnight at room temperature in the dark. The reaction was monitored by LCMS and once gone to completion, the mixture was diluted with CH₂Cl₂ and the organic was washed with water and brine. The organic was dried over Na₂SO₄ and the solvent was removed. The crude solid was purified by flash chromatography (100 % EtOAc) to afford a yellow solid (53 mg, 13 %).

¹H NMR (500 MHz, CDCl₃) δ 8.32 (d, J = 7.9 Hz, 1H), 8.24 – 8.19 (m, 2H), 7.97 (d, J = 8.6 Hz, 1H), 7.95 – 7.86 (m, 1H), 7.73 – 7.61 (m, 1H), 7.41 (d, J = 8.5 Hz, 2H), 4.97 (t, J = 4.9 Hz, 2H), 3.97 (t, J = 5.1 Hz, 2H), 3.95 (s, 3H), 3.31 (s, 3H).

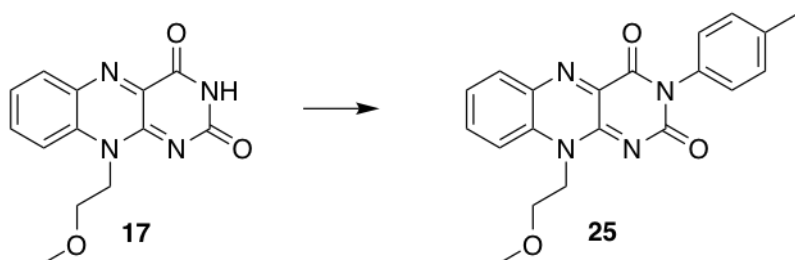
¹³C NMR (126 MHz, CDCl₃) δ 166.4, 154.8, 139.6, 136.1, 135.9, 133.0, 130.8, 130.4, 130.2, 128.5, 127.3, 126.8, 117.0, 69.5, 59.3, 52.3, 45.9.

IR ν_{max} (film)/cm⁻¹: 3431, 2097, 1716, 1667, 1557, 1282, 1112,

LCMS m/z (ESI) [M+H]⁺: calcd for [C₂₁H₁₈N₄O₅]⁺ 407.13, found 407.1

Mp >220 °C

10-(2-Methoxyethyl)-3-(p-tolyl)benzo[g]pteridine-2,4(3H,10H)-dione (25)



To a solution of **17** (136 mg, 0.5 mmol) in anhydrous DMF (12.5 mL) was added 4-tolylboronic acid (204 mg, 1.5 mmol, 3 equiv), Cu(OAc)₂ (182 mg, 1 mmol, 2 equiv), triethylamine (0.21 mL, 1.5 mmol, 3 equiv) and was stirred overnight at room temperature in the dark. The reaction was monitored by LCMS and once gone to completion, the mixture was diluted with CH₂Cl₂ and the organic was washed with water and brine. The organic was dried over Na₂SO₄ and the solvent was removed. The crude solid was purified by flash chromatography (100 % EtOAc) to afford a yellow solid (20 mg, 11 %).

¹H NMR (500 MHz, CDCl₃) δ 8.34 (s, 1H), 7.97 (s, 1H), 7.94 – 7.82 (m, 1H), 7.70 – 7.58 (m, 1H), 7.32 (t, J = 9.1 Hz, 2H), 7.20 (d, J = 8.0 Hz, 2H), 4.95 (t, J = 4.9 Hz, 2H), 3.96 (t, J = 4.9 Hz, 2H), 3.31 (s, 3H), 2.42 (s, 3H).

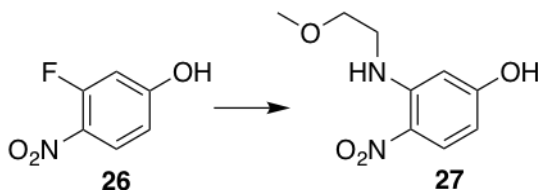
¹³C NMR (126 MHz, CDCl₃) δ 138.7, 135.6, 133.0, 132.8, 130.2, 127.8, 69.6, 59.3, 45.7, 21.4.

IR ν_{max} (film)/cm⁻¹: 3433, 2089, 1662, 1557,

LCMS m/z (ESI) [M+H]⁺: calcd for [C₂₀H₁₈N₄O₃]⁺ 363.14, found 363.2

Mp >220 °C

3-((2-Methoxyethyl)amino)-4-nitrophenol (27)



3-fluoro-4-nitrophenol (1 g, 6.4 mmol), Na₂CO₃ (340 mg, 3.2 mmol, 0.5 equiv) was dissolved in ethanol (2.5 mL) and heated to 80 °C. 2-Methoxyethylamine was added dropwise and the reaction was left at 80 °C overnight. The ethanol was removed under vacuum and H₂O (50 mL) was added. The product was extracted with 3x 50 mL CH₂Cl₂. The organic was dried over Na₂SO₄ and the solvent was removed to afford a dark orange solid (809 mg, 61 %).

¹H NMR (500 MHz, CDCl₃) δ 8.12 (d, *J* = 9.3 Hz, 1H), 6.23 (d, *J* = 2.4 Hz, 1H), 6.17 (dd, *J* = 9.3, 2.4 Hz, 1H), 3.68 (t, *J* = 5.4 Hz, 2H), 3.47 – 3.40 (m, 5H).

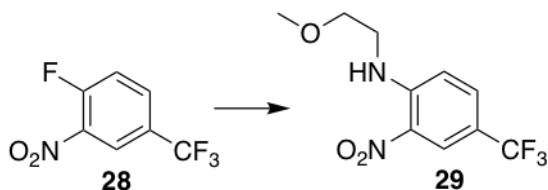
¹³C NMR (126 MHz, CDCl₃) δ 162.64, 129.82, 105.59, 97.82, 70.31, 59.12, 42.99.

LCMS *m/z* (ESI) [M-H]⁻: calcd for [C₉H₁₂N₂O₄]⁻ 211.08, found 211.2

IR ν_{max} (film)/cm⁻¹: 3212, 1578, 1166

Mp >220 °C

***N*-(2-methoxyethyl)-2-nitro-4-(trifluoromethyl)aniline (29)**



4-fluoro-3-nitrobenzotrifluoride (1 g, 4.8 mmol), Na₂CO₃ (254 mg, 2.4 mmol, 0.5 equiv) was dissolved in ethanol (2 mL) and heated to 80 °C. 2-methoxyethylamine (0.46 mL, 5.3 mmol, 2.2 equiv) was added dropwise and the reaction was left at 80 °C overnight. The ethanol was removed under vacuum and H₂O (50 mL) was added. The product was extracted with 3x 50 mL CH₂Cl₂. The organic was dried over Na₂SO₄ and the solvent was removed to afford a yellow solid (1.1 g, 86 %).

¹H NMR (400 MHz, CDCl₃) δ 8.47 (t, *J* = 5.8 Hz, 1H), 7.62 (dd, *J* = 9.1, 2.1 Hz, 1H), 6.96 (d, *J* = 9.0 Hz, 1H), 3.69 (t, *J* = 5.3 Hz, 2H), 3.58 – 3.50 (m, 2H), 3.44 (s, 3H).

¹³C NMR (126 MHz, DMSO) δ 147.4, 132.4, 132.4, 130.5, 124.5, 124.47 (q, *J* CF = 4.4 Hz), 116.7, 70.4, 58.6, 42.6.

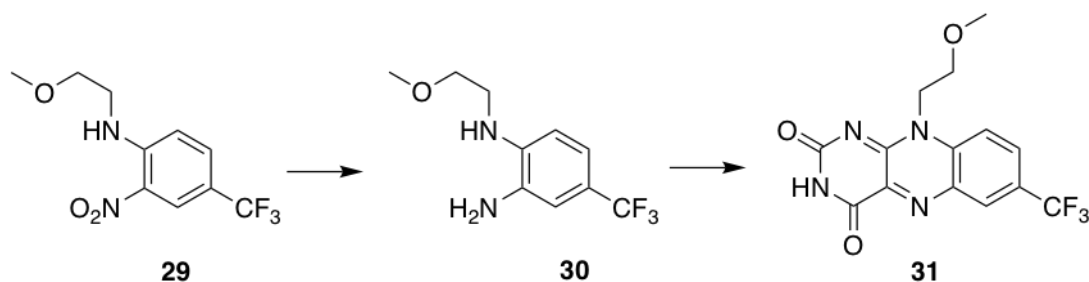
¹⁹F NMR (376 MHz, DMSO) δ -60.39.

IR ν_{max} (film)/cm⁻¹: 3379, 1571, 1247, 1111, 1016

LCMS *m/z* (ESI) [M+H]⁺: calcd for [C₁₀H₁₁F₃N₂O₃]⁺ 265.07, found 265.2

Mp >220 °C

10-(2-Methoxyethyl)-7-(trifluoromethyl)benzo[g]pteridine-2,4(3H,10H)-dione (31)



Compound **29** (1 g, 3.8 mmol) as dissolved in acetic acid (15 mL) and the mixture was degassed. Pd/C (12 mg, 0.11 mmol, 3 %) was added and the mixture was further degassed. The reaction vessel was placed under vacuum and purged with hydrogen. The reaction was left under hydrogen overnight at room temperature. LCMS showed the formation of compound **30** and the mixture was filtered through a pad of celite. Boric acid (1.17 g, 19.0 mmol, 5 equiv) and alloxan monohydrate (1.18 g, 7.4 mmol, 2 equiv) was added and stirred overnight at room temperature. Acetic acid was removed under vacuum and 50 mL H₂O was added. The product was extracted with 3x 100 mL CH₂Cl₂. The product precipitated and the DCM was filtered to collect a crude yellow solid. The crude solid was purified by column chromatography (100 % EtOAc) to afford a yellow solid (117 mg, 9 %)

¹H NMR (400 MHz, CDCl₃) δ 8.57 (s, 1H), 8.50 (s, 1H), 8.05 (dt, J = 9.4, 5.6 Hz, 2H), 4.91 (t, J = 6.2 Hz 2H), 3.94 (t, J = 6.2 Hz, 2H), 3.28 (s, 3H).

¹³C NMR (126 MHz, DMSO) δ 159.8, 155.9, 151.6, 141.1, 136.2, 134.2, 130.2, 129.1, 119.3, 68.9, 58.9, 45.0. *CF coupling was not observed in the spectra.

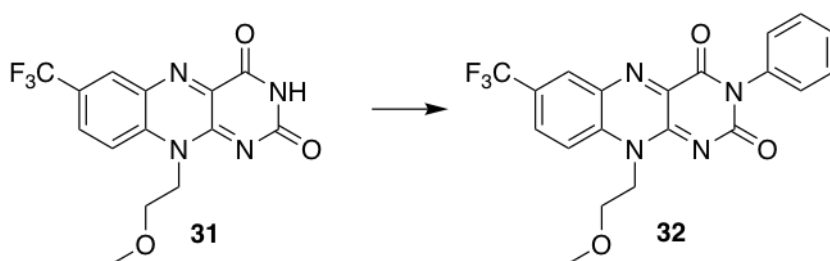
¹⁹F NMR (376 MHz, DMSO) δ -60.73.

IR ν_{max} (film)/cm⁻¹: 3415, 1709, 1164

LCMS m/z (ESI) [M+H]⁺: calcd for [C₁₄H₁₁F₃N₄O₃]⁺ 341.08, found 341.1

Mp >220 °C

10-(2-Methoxyethyl)-3-phenyl-7-(trifluoromethyl)benzo[g]pteridine-2,4(3H,10H)-dione
(32)



To a solution of **31** (100 mg, 0.3 mmol) in anhydrous DMF (3.75 mL) was added benzene boronic acid (110 mg, 0.9 mmol, 3 equiv), Cu(OAc)₂ (109 mg, 0.6 mmol, 2 equiv), triethylamine (0.125 mL, 0.9 mmol, 3 equiv) and was stirred overnight at room temperature in the dark. The reaction was monitored by LCMS and once gone to completion, the mixture was diluted with CH₂Cl₂ and the organic was washed with water and brine. The organic was dried over Na₂SO₄ and the solvent was removed. The crude solid was purified by flash chromatography (100 % EtOAc) to afford a brown solid (80 mg, 64 %).

¹H NMR (400 MHz, CDCl₃) δ 8.63 (s, 1H), 8.11 (s, 1H), 8.03 (d, J = 8.9 Hz, 1H), 7.55 (t, J = 7.1 Hz, 2H), 7.47 (t, J = 7.4 Hz, 1H), 7.32 (d, J = 7.5 Hz, 2H), 4.94 (s, 2H), 3.97 (s, 2H), 3.29 (s, 3H).

¹³C NMR (126 MHz, DMSO) δ 159.5, 155.1, 150.6, 140.8, 136.9, 136.2, 134.4, 129.8, 129.5, 128.9, 128.6, 119.5, 115.7, 69.0, 59.0, 45.1, 28.4.

*CF coupling was not observed in the spectra.

¹⁹F NMR (376 MHz, DMSO) δ -60.72.

IR ν_{max} (film)/cm⁻¹: 3433, 2090, 1672, 1562, 1127

LCMS m/z (ESI) [M+H]⁺: calcd for [C₂₀H₁₅F₃N₄O₃]⁺ 417.11, found 417.1

Mp >220 °C

References

- Adams, J., Behnke, M., Chen, S., Cruickshank, A. A., Dick, L. R., Grenier, L., Klunder, J. M., Ma, Y.-T., Plamondon, L., and Stein, R. L. (1998). Potent and selective inhibitors of the proteasome: dipeptidyl boronic acids. *Bioorganic & medicinal chemistry letters*, 8(4):333–338.
- Arias, H. R., Gu, R.-X., Feuerbach, D., Guo, B.-B., Ye, Y., and Wei, D.-Q. (2011). Novel positive allosteric modulators of the human $\alpha 7$ nicotinic acetylcholine receptor. *Biochemistry*, 50(23):5263–5278.
- Aronchik, I., Kundu, A., Quirit, J. G., and Firestone, G. L. (2014). The antiproliferative response of indole-3-carbinol in human melanoma cells is triggered by an interaction with nedd4-1 and disruption of wild-type pten degradation. *Molecular Cancer Research*, 12(11):1621–1634.
- Attenberger, B., Schmaderer, H., and Koenig, B. (2008). Copper-mediated 3-n-arylation of flavins. *Synthesis*, 2008(11):1767–1774.
- Bentley, M. L., Corn, J. E., Dong, K. C., Phung, Q., Cheung, T. K., and Cochran, A. G. (2011). Recognition of ubch5c and the nucleosome by the bmi1/ring1b ubiquitin ligase complex. *The EMBO journal*, 30(16):3285–3297.
- Bernassola, F., Karin, M., Ciechanover, A., and Melino, G. (2008). The hect family of e3 ubiquitin ligases: multiple players in cancer development. *Cancer cell*, 14(1):10–21.
- Bogyo, M., McMaster, J. S., Gaczynska, M., Tortorella, D., Goldberg, A. L., and Ploegh, H. (1997). Covalent modification of the active site threonine of proteasomal β subunits and the escherichia coli homolog hslv by a new class of inhibitors. *Proceedings of the National Academy of Sciences*, 94(13):6629–6634.
- Brylinski, M. (2018). Aromatic interactions at the ligand–protein interface: Implications for the development of docking scoring functions. *Chemical biology & drug design*, 91(2):380–390.
- Brzovic, P. S., Rajagopal, P., Hoyt, D. W., King, M.-C., and Klevit, R. E. (2001). Structure of a brca1–bard1 heterodimeric ring–ring complex. *Nature Structural & Molecular Biology*, 8(10):833.
- Buckley, D. L., Van Molle, I., Gareiss, P. C., Tae, H. S., Michel, J., Noblin, D. J., Jorgensen, W. L., Ciulli, A., and Crews, C. M. (2012). Targeting the von hippel–lindau e3 ubiquitin ligase using small molecules to disrupt the vhl/hif-1 α interaction. *Journal of the American Chemical Society*, 134(10):4465–4468.
- Bulman Page, P. C., Goodyear, R. L., Horton, A. E., Chan, Y., Karim, R., O’Connell, M. A., Hamilton, C., Slawin, A. M., Buckley, B. R., and Allin, S. M. (2017). Formal total synthesis of (+)-c9-deoxyomuralide from l-leucine using a double sacrificial chirality transfer approach. *The Journal of organic chemistry*, 82(23):12209–12223.

- Burgess, A., Chia, K. M., Haupt, S., Thomas, D., Haupt, Y., and Lim, E. (2016). Clinical overview of mdm2/x-targeted therapies. *Frontiers in oncology*, 6:7.
- Burslem, G. M. and Crews, C. M. (2017). Small-molecule modulation of protein homeostasis. *Chemical reviews*, 117(17):11269–11301.
- Cao, Y., Wang, C., Zhang, X., Xing, G., Lu, K., Gu, Y., He, F., and Zhang, L. (2014). Selective small molecule compounds increase bmp-2 responsiveness by inhibiting smurf1-mediated smad1/5 degradation. *Scientific reports*, 4:4965.
- Ceccarelli, D. F., Tang, X., Pelletier, B., Orlicky, S., Xie, W., Plantevin, V., Neculai, D., Chou, Y.-C., Ogunjimi, A., Al-Hakim, A., et al. (2011). An allosteric inhibitor of the human cdc34 ubiquitin-conjugating enzyme. *Cell*, 145(7):1075–1087.
- Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., Jacobsen, A., Byrne, C. J., Heuer, M. L., Larsson, E., et al. (2012). The cbio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data.
- Chantry, A. (2011). Wwp2 ubiquitin ligase and its isoforms: new biological insight and promising disease targets. *Cell cycle*, 10(15):2437–2439.
- Chen, D., Gehringer, M., and Lorenz, S. (2018). Developing small-molecule inhibitors of hect-type ubiquitin ligases for therapeutic applications: Challenges and opportunities. *ChemBioChem*, 19(20):2123–2135.
- Chen, H. I. and Sudol, M. (1995). The ww domain of yes-associated protein binds a proline-rich ligand that differs from the consensus established for src homology 3-binding modules. *Proceedings of the National Academy of Sciences*, 92(17):7819–7823.
- Chen, Z., Jiang, H., Xu, W., Li, X., Dempsey, D. R., Zhang, X., Devreotes, P., Wolberger, C., Amzel, L. M., Gabelli, S. B., et al. (2017). A tunable brake for hect ubiquitin ligases. *Molecular cell*, 66(3):345–357.
- Cheng, K., Wang, X., Zhang, S., and Yin, H. (2012). Discovery of small-molecule inhibitors of the tlr1/tlr2 complex. *Angewandte Chemie International Edition*, 51(49):12246–12249.
- Ciccone, L., Vera, L., Tepshi, L., Rosalia, L., Rossello, A., and Stura, E. A. (2015). Multi-component mixtures for cryoprotection and ligand solubilization. *Biotechnology Reports*, 7:120–127.
- Ciehanover, A., Hod, Y., and Hershko, A. (1978). A heat-stable polypeptide component of an atp-dependent proteolytic system from reticulocytes. *Biochemical and biophysical research communications*, 81(4):1100–1105.
- Cooper, G. M., Hausman, R. E., and Hausman, R. E. (2000). *The cell: a molecular approach*, volume 10. ASM press Washington, DC.
- Crews, C. M. (2010). Targeting the undruggable proteome: the small molecules of my dreams. *Chemistry & biology*, 17(6):551–555.
- Dahlin, J. L., Nissink, J. W. M., Strasser, J. M., Francis, S., Higgins, L., Zhou, H., Zhang, Z., and Walters, M. A. (2015). Pains in the assay: chemical mechanisms of assay interference and promiscuous enzymatic inhibition observed during a sulfhydryl-scavenging hts. *Journal of medicinal chemistry*, 58(5):2091–2113.

- Darvesh, S., McDonald, R. S., Darvesh, K. V., Mataija, D., Mothana, S., Cook, H., Carneiro, K. M., Richard, N., Walsh, R., and Martin, E. (2006). On the active site for hydrolysis of aryl amides and choline esters by human cholinesterases. *Bioorganic & medicinal chemistry*, 14(13):4586–4599.
- Dou, H., Buetow, L., Sibbet, G. J., Cameron, K., and Huang, D. T. (2012). Birc7–e2 ubiquitin conjugate structure reveals the mechanism of ubiquitin transfer by a ring dimer. *Nature structural & molecular biology*, 19(9):876.
- Epple, R., Wallenborn, E.-U., and Carell, T. (1997). Investigation of flavin-containing dna-repair model compounds. *Journal of the American Chemical Society*, 119(32):7440–7451.
- Evans, P. and McCoy, A. (2008). An introduction to molecular replacement. *Acta Crystallographica Section D: Biological Crystallography*, 64(1):1–10.
- Fenteany, G., Standaert, R. F., Lane, W. S., Choi, S., Corey, E. J., and Schreiber, S. L. (1995). Inhibition of proteasome activities and subunit-specific amino-terminal threonine modification by lactacystin. *Science*, 268(5211):726–731.
- Galdeano, C., Gadd, M. S., Soares, P., Scaffidi, S., Van Molle, I., Birced, I., Hewitt, S., Dias, D. M., and Ciulli, A. (2014). Structure-guided design and optimization of small molecules targeting the protein–protein interaction between the von hippel–lindau (vhl) e3 ubiquitin ligase and the hypoxia inducible factor (hif) alpha subunit with in vitro nanomolar affinities. *Journal of medicinal chemistry*, 57(20):8657–8663.
- Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cbiportal. *Sci. Signal.*, 6(269):pl1–pl1.
- Garcia-Ruiz, J. M. (2003). Nucleation of protein crystals. *Journal of Structural Biology*, 142(1):22–31.
- Goldstein, G., Scheid, M., Hammerling, U., Schlesinger, D., Niall, H., and Boyse, E. (1975). Isolation of a polypeptide that has lymphocyte-differentiating properties and is probably represented universally in living cells. *Proceedings of the National Academy of Sciences*, 72(1):11–15.
- Gong, W., Zhang, X., Zhang, W., Li, J., and Li, Z. (2015). Structure of the hect domain of human wwp2. *Acta Crystallographica Section F: Structural Biology Communications*, 71(10):1251–1257.
- Groen, E. J. and Gillingwater, T. H. (2015). Uba1: at the crossroads of ubiquitin homeostasis and neurodegeneration. *Trends in Molecular Medicine*, 21(10):622–632.
- Groettrup, M., Pelzer, C., Schmidtke, G., and Hofmann, K. (2008). Activating the ubiquitin family: Uba6 challenges the field. *Trends in biochemical sciences*, 33(5):230–237.
- Gulder, T. A. and Moore, B. S. (2010). Salinosporamide natural products: Potent 20 s proteasome inhibitors as promising cancer chemotherapeutics. *Angewandte Chemie International Edition*, 49(49):9346–9367.
- Halgren, T. (2007). New method for fast and accurate binding-site identification and analysis. *Chemical biology & drug design*, 69(2):146–148.
- Halgren, T. A. (2009). Identifying and characterizing binding sites and assessing druggability. *Journal of chemical information and modeling*, 49(2):377–389.

- Hansen, T., Rossi, M., Roperch, J., Ansell, K., Simpson, K., Taylor, D., Mathon, N., Knight, R., and Melino, G. (2007). Itch inhibition regulates chemosensitivity in vitro. *Biochemical and biophysical research communications*, 361(1):33–36.
- Hart, D. J. and Huang, H. C. (1988). Total synthesis of (.-)-pleurotin and (.-)-dihydropleurotin acid. *Journal of the American Chemical Society*, 110(5):1634–1635.
- Hershko, A., Ciechanover, A., Heller, H., Haas, A. L., and Rose, I. A. (1980). Proposed role of atp in protein breakdown: conjugation of protein with multiple chains of the polypeptide of atp-dependent proteolysis. *Proceedings of the National Academy of Sciences*, 77(4):1783–1786.
- Hershko, A., Ciechanover, A., and Rose, I. A. (1979). Resolution of the atp-dependent proteolytic system from reticulocytes: a component that interacts with atp. *Proceedings of the National Academy of Sciences*, 76(7):3107–3110.
- Huang, L., Kinnucan, E., Wang, G., Beaudenon, S., Howley, P. M., Huibregtse, J. M., and Pavletich, N. P. (1999). Structure of an e6ap-ubch7 complex: insights into ubiquitination by the e2-e3 enzyme cascade. *Science*, 286(5443):1321–1326.
- Jacobsen, N. E. (2007). *NMR spectroscopy explained: simplified theory, applications and examples for organic chemistry and structural biology*. John Wiley & Sons.
- Jasial, S., Hu, Y., and Bajorath, J. (2017). How frequently are pan-assay interference compounds active? large-scale analysis of screening data reveals diverse activity profiles, low global hit frequency, and many consistently inactive compounds. *Journal of medicinal chemistry*, 60(9):3879–3886.
- Kamadurai, H. B., Souphron, J., Scott, D. C., Duda, D. M., Miller, D. J., Stringer, D., Piper, R. C., and Schulman, B. A. (2009). Insights into ubiquitin transfer cascades from a structure of a ubch5b ubiquitin-ubtndd4l complex. *Molecular cell*, 36(6):1095–1102.
- Kathman, S. G., Span, I., Smith, A. T., Xu, Z., Zhan, J., Rosenzweig, A. C., and Statsyuk, A. V. (2015). A small molecule that switches a ubiquitin ligase from a processive to a distributive enzymatic mechanism. *Journal of the American Chemical Society*, 137(39):12442–12445.
- Kathman, S. G., Xu, Z., and Statsyuk, A. V. (2014). A fragment-based method to discover irreversible covalent inhibitors of cysteine proteases. *Journal of medicinal chemistry*, 57(11):4969–4974.
- Kitagaki, J., Yang, Y., Saavedra, J., Colburn, N., Keefer, L., and Perantoni, A. (2009). Nitric oxide prodrug js-k inhibits ubiquitin e1 and kills tumor cells retaining wild-type p53. *Oncogene*, 28(4):619.
- Kuo, Y.-H. and Shih, K.-S. (1991). A new method for preparation of 3-hydroxypyridines from furfurylamines by photooxygenation. *Chemical and pharmaceutical bulletin*, 39(1):181–183.
- Lacroix, A., Schabat, D., Clerin, D., and Fleury, J.-P. (1987). Nouvelle voie d'accès aux iso-alloxazines par l'intermédiaire d'alkyl-1-amino-2-tétrahydro-1, 5, 6, 7-quinoline-3-carbonitriles (ou carboxylate de méthyle). *Bulletin de la Société chimique de France*, (6):1065–1072.
- Lee, I. and Schindelin, H. (2008). Structural insights into e1-catalyzed ubiquitin activation and transfer to conjugating enzymes. *Cell*, 134(2):268–278.

- Lee, Y.-R., Chen, M., Lee, J. D., Zhang, J., Lin, S.-Y., Fu, T.-M., Chen, H., Ishikawa, T., Chiang, S.-Y., Katon, J., et al. (2019). Reactivation of pten tumor suppressor for cancer treatment through inhibition of a myc-wwp1 inhibitory pathway. *Science*, 364(6441):eaau0159.
- Liu, X., Sun, L., Gursel, D. B., Cheng, C., Huang, S., Rademaker, A. W., Khan, S. A., Yin, J., and Kiyokawa, H. (2017). The non-canonical ubiquitin activating enzyme uba6 suppresses epithelial-mesenchymal transition of mammary epithelial cells. *Oncotarget*, 8(50):87480.
- Lorenz, S. (2018). Structural mechanisms of hect-type ubiquitin ligases. *Biological chemistry*, 399(2):127–145.
- Mace, P. D., Linke, K., Feltham, R., Schumacher, F.-R., Smith, C. A., Vaux, D. L., Silke, J., and Day, C. L. (2008). Structures of the ciap2 ring domain reveal conformational changes associated with ubiquitin-conjugating enzyme (e2) recruitment. *Journal of Biological Chemistry*, 283(46):31633–31640.
- Maddika, S., Kavela, S., Rani, N., Palicharla, V. R., Pokorny, J. L., Sarkaria, J. N., and Chen, J. (2011). Wwp2 is an e3 ubiquitin ligase for pten. *Nature cell biology*, 13(6):728.
- Mannhold, R., Kubinyi, H., and Folkers, G. (2015). *Fragment-based drug discovery: lessons and outlook*. John Wiley & Sons.
- Mari, S., Ruetalo, N., Maspero, E., Stoffregen, M. C., Pasqualato, S., Polo, S., and Wiesner, S. (2014). Structural and functional framework for the autoinhibition of nedd4-family ubiquitin ligases. *Structure*, 22(11):1639–1649.
- Maspero, E., Mari, S., Valentini, E., Musacchio, A., Fish, A., Pasqualato, S., and Polo, S. (2011). Structure of the hect: ubiquitin complex and its role in ubiquitin chain elongation. *EMBO reports*, 12(4):342–349.
- Mayer, M. and Meyer, B. (2001). Group epitope mapping by saturation transfer difference nmr to identify segments of a ligand in direct contact with a protein receptor. *Journal of the American Chemical Society*, 123(25):6108–6117.
- McNicholas, S., Potterton, E., Wilson, K., and Noble, M. (2011). Presenting your structures: the ccp4mg molecular-graphics software. *Acta Crystallographica Section D: Biological Crystallography*, 67(4):386–394.
- McPherson, A. (2004). Introduction to protein crystallization. *Methods*, 34(3):254–265.
- Meng, L., Mohan, R., Kwok, B. H., Elofsson, M., Sin, N., and Crews, C. M. (1999). Epoxomicin, a potent and selective proteasome inhibitor, exhibits in vivo antiinflammatory activity. *Proceedings of the National Academy of Sciences*, 96(18):10403–10408.
- Metzger, M. B., Pruneda, J. N., Klevit, R. E., and Weissman, A. M. (2014). Ring-type e3 ligases: master manipulators of e2 ubiquitin-conjugating enzymes and ubiquitination. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1843(1):47–60.
- Monaco, S., Tailford, L. E., Juge, N., and Angulo, J. (2017). Differential epitope mapping by std nmr spectroscopy to reveal the nature of protein–ligand contacts. *Angewandte Chemie International Edition*, 56(48):15289–15293.
- Mund, T., Lewis, M. J., Maslen, S., and Pelham, H. R. (2014). Peptide and small molecule inhibitors of hect-type ubiquitin ligases. *Proceedings of the National Academy of Sciences*, 111(47):16736–16741.

- Niesen, F. H., Berglund, H., and Vedadi, M. (2007). The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature protocols*, 2(9):2212.
- Olsen, S. K. and Lima, C. D. (2013). Structure of a ubiquitin e1-e2 complex: insights to e1-e2 thioester transfer. *Molecular cell*, 49(5):884–896.
- Patching, S. G. (2014). Surface plasmon resonance spectroscopy for characterisation of membrane protein–ligand interactions and its potential for drug discovery. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1838(1):43–55.
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004). Ucsf chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*, 25(13):1605–1612.
- Pickart, C. M. and Eddins, M. J. (2004). Ubiquitin: structures, functions, mechanisms. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1695(1-3):55–72.
- Pirozzi, G., McConnell, S. J., Uveges, A. J., Carter, J. M., Sparks, A. B., Kay, B. K., and Fowlkes, D. M. (1997). Identification of novel human ww domain-containing proteins by cloning of ligand targets. *Journal of Biological Chemistry*, 272(23):14611–14616.
- Plant, P. J., Yeger, H., Staub, O., Howard, P., and Rotin, D. (1997). The c2 domain of the ubiquitin protein ligase nedd4 mediates ca²⁺-dependent plasma membrane localization. *Journal of Biological Chemistry*, 272(51):32329–32336.
- Plechanovová, A., Jaffray, E. G., Tatham, M. H., Naismith, J. H., and Hay, R. T. (2012). Structure of a ring e3 ligase and ubiquitin-loaded e2 primed for catalysis. *Nature*, 489(7414):115.
- Quirit, J. G., Lavrenov, S. N., Poindexter, K., Xu, J., Kyauk, C., Durkin, K. A., Aronchik, I., Tomasiak, T., Solomatin, Y. A., Preobrazhenskaya, M. N., et al. (2017). Indole-3-carbinol (i3c) analogues are potent small molecule inhibitors of nedd4-1 ubiquitin ligase activity that disrupt proliferation of human melanoma cells. *Biochemical pharmacology*, 127:13–27.
- Raoof, A., Depledge, P., Hamilton, N. M., Hamilton, N. S., Hitchin, J. R., Hopkins, G. V., Jordan, A. M., Maguire, L. A., McGonagle, A. E., Mould, D. P., et al. (2013). Toxoflavins and deazaflavins as the first reported selective small molecule inhibitors of tyrosyl-dna phosphodiesterase ii. *Journal of medicinal chemistry*, 56(16):6352–6370.
- Rizo, J. and Südhof, T. C. (1998). C2-domains, structure and function of a universal ca²⁺-binding domain. *Journal of Biological Chemistry*, 273(26):15879–15882.
- Ronchi, V. P., Klein, J. M., and Haas, A. L. (2013). E6ap/ube3a ubiquitin ligase harbors two e2 ubiquitin binding sites. *Journal of Biological Chemistry*, 288(15):10349–10360.
- Rossi, M., Rotblat, B., Ansell, K., Amelio, I., Caraglia, M., Misso, G., Bernassola, F., Cavasotto, C. N., Knight, R., Ciechanover, A., et al. (2014). High throughput screening for inhibitors of the hec ubiquitin e3 ligase itch identifies antidepressant drugs as regulators of autophagy. *Cell death & disease*, 5(5):e1203.
- Sakamoto, K. M., Kim, K. B., Kumagai, A., Mercurio, F., Crews, C. M., and Deshaies, R. J. (2001). Protacs: chimeric molecules that target proteins to the skp1–cullin–f box complex for ubiquitination and degradation. *Proceedings of the National Academy of Sciences*, 98(15):8554–8559.

- Sakata, E., Satoh, T., Yamamoto, S., Yamaguchi, Y., Yagi-Utsumi, M., Kurimoto, E., Tanaka, K., Wakatsuki, S., and Kato, K. (2010). Crystal structure of ubch5b ubiquitin intermediate: insight into the formation of the self-assembled e2 ub conjugates. *Structure*, 18(1):138–147.
- Sanarico, A., Ronchini, C., Croce, A., Memmi, E., Cammarata, U., De Antoni, A., Lavorgna, S., Divona, M., Giacò, L., Melloni, G., et al. (2018). The e3 ubiquitin ligase wwp1 sustains the growth of acute myeloid leukaemia. *Leukemia*, 32(4):911.
- Sánchez-Tena, S., Cubillos-Rojas, M., Schneider, T., and Rosa, J. L. (2016). Functional and pathological relevance of herc family proteins: a decade later. *Cellular and Molecular Life Sciences*, 73(10):1955–1968.
- Schäfer, A., Kuhn, M., and Schindelin, H. (2014). Structure of the ubiquitin-activating enzyme loaded with two ubiquitin molecules. *Acta Crystallographica Section D: Biological Crystallography*, 70(5):1311–1320.
- Scheffner, M. and Kumar, S. (2014). Mammalian hect ubiquitin-protein ligases: biological and pathophysiological aspects. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1843(1):61–74.
- Schulman, B. A. and Harper, J. W. (2009). Ubiquitin-like protein activation by e1 enzymes: the apex for downstream signalling pathways. *Nature reviews Molecular cell biology*, 10(5):319.
- Soares, P., Gadd, M. S., Frost, J., Galdeano, C., Ellis, L., Epemolu, O., Rocha, S., Read, K. D., and Ciulli, A. (2017). Group-based optimization of potent and cell-active inhibitors of the von hippel–lindau (vhl) e3 ubiquitin ligase: Structure–activity relationships leading to the chemical probe (2 s, 4 r)-1-((s)-2-(1-cyanocyclopropanecarboxamido)-3, 3-dimethylbutanoyl)-4-hydroxy-n-(4-(4-methylthiazol-5-yl) benzyl) pyrrolidine-2-carboxamide (vh298). *Journal of medicinal chemistry*, 61(2):599–618.
- Soond, S. M., Smith, P. G., Wahl, L., Swingler, T. E., Clark, I. M., Hemmings, A. M., and Chantry, A. (2013). Novel wwp2 ubiquitin ligase isoforms as potential prognostic markers and molecular targets in cancer. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1832(12):2127–2135.
- Stewart, M. D., Ritterhoff, T., Klevit, R. E., and Brzovic, P. S. (2016). E2 enzymes: more than just middle men. *Cell research*, 26(4):423.
- Strickson, S., Campbell, D. G., Emmerich, C. H., Knebel, A., Plater, L., Ritorto, M. S., Shpiro, N., and Cohen, P. (2013). The anti-inflammatory drug bay 11-7082 suppresses the myd88-dependent signalling network by targeting the ubiquitin system. *Biochemical Journal*, 451(3):427–437.
- Sun, L. and Chen, Z. J. (2004). The novel functions of ubiquitination in signaling. *Current opinion in cell biology*, 16(2):119–126.
- Taylor, D. (1959). 555. some experiments with furans. *Journal of the Chemical Society (Resumed)*, pages 2767–2769.
- Taylor, G. (2003). The phase problem. *Acta Crystallographica Section D: Biological Crystallography*, 59(11):1881–1890.
- Teicher, B. A. and Tomaszewski, J. E. (2015). Proteasome inhibitors. *Biochemical pharmacology*, 96(1):1–9.

- Van Driessche, A. E., Van Gerven, N., Bomans, P. H., Joosten, R. R., Friedrich, H., Gil-Carton, D., Sommerdijk, N. A., and Sleutel, M. (2018). Molecular nucleation mechanisms and control strategies for crystal polymorph selection. *Nature*, 556(7699):89.
- Van Veldhoven, J., Blad, C., Artsen, C., Klopman, C., Wolfram, D., Abdelkadir, M., Lane, J., Brussee, J., and Ijzerman, A. (2011). Structure–activity relationships of trans-substituted-propenoic acid derivatives on the nicotinic acid receptor hca2 (gpr109a). *Bioorganic & medicinal chemistry letters*, 21(9):2736–2739.
- Vantourout, J. C., Miras, H. N., Isidro-Llobet, A., Sproules, S., and Watson, A. J. (2017). Spectroscopic studies of the chan–lam amination: a mechanism-inspired solution to boronic ester reactivity. *Journal of the American Chemical Society*, 139(13):4769–4779.
- Verdecia, M. A., Joazeiro, C. A., Wells, N. J., Ferrer, J.-L., Bowman, M. E., Hunter, T., and Noel, J. P. (2003). Conformational flexibility underlies ubiquitin ligation mediated by the wwp1 hect domain e3 ligase. *Molecular cell*, 11(1):249–259.
- Viegas, A., Manso, J., Nobrega, F. L., and Cabrita, E. J. (2011). Saturation-transfer difference (std) nmr: a simple and fast method for ligand screening and characterization of protein binding. *Journal of chemical Education*, 88(7):990–994.
- Vijay-Kumar, S., Bugg, C. E., and Cook, W. J. (1987). Structure of ubiquitin refined at 1.8 Å resolution. *Journal of molecular biology*, 194(3):531–544.
- Vinitzky, A., Michaud, C., Powers, J. C., and Orłowski, M. (1992). Inhibition of the chymotrypsin-like activity of the pituitary multicatalytic proteinase complex. *Biochemistry*, 31(39):9421–9428.
- Walpole, S., Monaco, S., Nepravishta, R., and Angulo, J. (2019). Std nmr as a technique for ligand screening and structural studies. *Methods Enzymol*, 617:423–451.
- Wang, Y., Jiang, X., Feng, F., Liu, W., and Sun, H. (2019). Degradation of protein by protac and other strategies. *Acta Pharmaceutica Sinica B*.
- Watt, J. (2018). *Development of novel WWP2 ubiquitin ligase inhibitors using structural and biochemical approaches*. PhD thesis, University of East Anglia.
- Watt, J. E., Hughes, G. R., Walpole, S., Monaco, S., Stephenson, G. R., Bulman Page, P. C., Hemmings, A. M., Angulo, J., and Chantry, A. (2018). Discovery of small molecule wwp2 ubiquitin ligase inhibitors. *Chemistry–A European Journal*, 24(67):17677–17680.
- Whitfield, H., Gilmartin, M., Baker, K., Riley, A. M., Godage, H. Y., Potter, B. V., Hemmings, A. M., and Brearley, C. A. (2018). A fluorescent probe identifies active site ligands of inositol pentakisphosphate 2-kinase. *Journal of medicinal chemistry*, 61(19):8838–8846.
- Wiesner, S., Ogunjimi, A. A., Wang, H.-R., Rotin, D., Sicheri, F., Wrana, J. L., and Forman-Kay, J. D. (2007). Autoinhibition of the hect-type ubiquitin ligase smurf2 through its c2 domain. *Cell*, 130(4):651–662.
- Windheim, M., Peggie, M., and Cohen, P. (2008). Two different classes of e2 ubiquitin-conjugating enzymes are required for the mono-ubiquitination of proteins and elongation by polyubiquitin chains with a specific topology. *Biochemical Journal*, 409(3):723–729.
- Yang, X., Brownell, J. E., Xu, Q., Zhu, F., Ma, J., Loke, H.-K., Rollins, N., Soucy, T. A., Minissale, J. J., Thomas, M. P., et al. (2013). Absolute quantification of e1, ubiquitin-like proteins and nedd8–mln4924 adduct by mass spectrometry. *Cell biochemistry and biophysics*, 67(1):139–147.

- Yang, Y., Kitagaki, J., Dai, R.-M., Tsai, Y. C., Lorick, K. L., Ludwig, R. L., Pierre, S. A., Jensen, J. P., Davydov, I. V., Oberoi, P., et al. (2007). Inhibitors of ubiquitin-activating enzyme (e1), a new class of potential cancer therapeutics. *Cancer research*, 67(19):9472–9481.
- Yao, P.-H. E., Kumar, S., Liu, Y.-L., Fang, C.-P., Liu, C.-C., and Sun, C.-M. (2017). Diversity-oriented synthesis of coumarin-linked benzimidazoles via a one-pot, three-step, intramolecular knoevenagel cyclization. *ACS combinatorial science*, 19(4):271–275.
- Yin, Q., Lin, S.-C., Lamothe, B., Lu, M., Lo, Y.-C., Hura, G., Zheng, L., Rich, R. L., Campos, A. D., Myszk, D. G., et al. (2009). E2 interaction and dimerization in the crystal structure of traf6. *Nature structural & molecular biology*, 16(6):658.
- Zhang, M., Windheim, M., Roe, S. M., Pegg, M., Cohen, P., Prodromou, C., and Pearl, L. H. (2005). Chaperoned ubiquitylation—crystal structures of the chip u box e3 ubiquitin ligase and a chip-ubc13-uev1a complex. *Molecular cell*, 20(4):525–538.
- Zhang, W., Wu, K.-P., Sartori, M. A., Kamadurai, H. B., Ordureau, A., Jiang, C., Mercredi, P. Y., Murchie, R., Hu, J., Persaud, A., et al. (2016). System-wide modulation of hec e3 ligases with selective ubiquitin variant probes. *Molecular cell*, 62(1):121–136.
- Zheng, N., Wang, P., Jeffrey, P. D., and Pavletich, N. P. (2000). Structure of a c-cbl–ubch7 complex: Ring domain function in ubiquitin-protein ligases. *Cell*, 102(4):533–539.
- Zou, X., Levy-Cohen, G., and Blank, M. (2015). Molecular functions of nedd4 e3 ubiquitin ligases in cancer. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, 1856(1):91–106.